UNIVERSIDADE FEDERAL DO PARANÁ

CAROLINA MATHIAS

IDENTIFICAÇÃO DE lncRNAs COMO REGULADORES DE REDES TRANSCRICIONAIS NO CÂNCER DE MAMA

CURITIBA

2021

CAROLINA MATHIAS

IDENTIFICAÇÃO DE lncRNAs COMO REGULADORES DE REDES TRANSCRICIONAIS NO CÂNCER DE MAMA

Tese de doutorado apresentada ao Programa de Pós-Graduação em Genética do Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, como requisito parcial para obtenção do título de Doutor em Genética.

Orientadora: Prof.ª Dr.ª Jaqueline Carvalho de Oliveira.
Co-orientadora: Prof.ª Dr.ª Daniela Fiori Gradia
Co-orientador: Prof. Dr. Mauro A. A. Castro

CURITIBA
2021

Universidade Federal do Paraná
Sistema de Bibliotecas
(Giana Mara Seniski Silva – CRB/9 1406)

Mathias, Carolina
Identificação de lncRNAs como reguladores de redes transcricionais no câncer de mama. / Carolina Mathias. – Curitiba, 2021.
84 p.: il.

Orientadora: Jaqueline Carvalho de Oliveira.
Coorientadores: Daniela Fiori Gradia e Mauro A. A. Castro.

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Mamas - Câncer. 2. Bioinformática. 3. Biomarcadores. I. Título. II. Oliveira, Jaqueline Carvalho de. III. Gradia, Daniela Fiori, 1971-. IV. Castro, Mauro Antônio Alves. V. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (22. ed.) 616.99449

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA - 40001016006P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **CAROLINA MATHIAS** intitulada: **Identificaço de lncRNAs reguladores de redes transcricionais no câncer de mama**, sob orientação da Profa. Dra. JAQUELINE CARVALHO DE OLIVEIRA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 06 de Agosto de 2021.

Assinatura Eletrônica
09/08/2021 14:07:32.0
JAQUELINE CARVALHO DE OLIVEIRA
Presidente da Banca Examinadora

Assinatura Eletrônica
09/08/2021 18:58:06.0
JULIA ALEJANDRA PEZUK
Avaliador Externo (UNIVERSIDADE ANHANGUERA DE SÃO PAULO, UNIAN/SP)

Assinatura Eletrônica
09/08/2021 15:12:20.0
MARIA LUIZA PETZL ERLER
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
09/08/2021 15:41:23.0
WILSON ARAUJO DA SILVA JUNIOR
Avaliador Externo (FACULDADE DE MEDICINA DE RIBEIRĀO PRETO - USP)

AGRADECIMENTOS

À minha orientadora Profª Dra. Jaqueline Carvalho de Oliveira, gostaria de agradecer imensamente por tudo. Pela confiança que depositou em mim para realizar este trabalho, por todo o conhecimento, científico, profissional e pessoal. Por ter sempre me apoiado, mesmo nos momentos em que parecia que tudo iria dar errado. Agradeço muito por ter tido a sorte de ser sua aluna. Tudo o que conquistei hoje é fruto dessa parceria e amizade que temos. Você é uma inspiração pra mim!

À minha co-orientadora Profª Dra. Daniela Fiori Gradia, obrigada por tudo nesses quatro anos. Obrigada por ser uma grande orientadora, pelas discussões científicas e pela amizade que cultivamos nestes anos. Sou muito grata por ter você na minha vida profissional e pessoal. Com certeza, é uma grande inspiração para mim.

Ao meu co-orientador Prof.Dr. Mauro Antônio Alves Castro por ter me apresentado ao mundo da bioinformática e ter me recebido em seu laboratório. Obrigada por tudo o que me ensinou e por ter acreditado em mim.

Aos professores do Laboratório de Citogenética Humana e Oncogenética, Prof.ª Dr.ª Enilze Maria de Souza Fonseca Ribeiro e Prof. Dr. Iglenir João Cavalli por terem me aceito como estudante de iniciação científica no ano de 2011 no laboratório. Aprendi muito com vocês nesses anos. E sou imensamente grata por esses 10 anos de aprendizado.

Ao meu supervisor Dr. George Adrian Calin, obrigada por me receber em seu laboratório no MD Anderson Cancer Center. Foram meses de grande crescimento científico que vou levar para o resto da minha vida científica.

Às agências de fomento CNPq e a CAPES, que possibilitaram o desenvolvimento deste trabalho.

A todos que convivi no Laboratório de Citogenética Humana e Oncogenética, durante esses 10 anos. Em especial, agradeço àquelas que além de parceiras de trabalho se tornaram grandes amigas. À Ana Flávia Kohler e à Vanessa Nascimento Kozak por sempre me proporcionarem momentos de muita risada, de compreensão e de incentivo.

Às amigas e parceiras de trabalho Danielle Malheiros Ferreira e Fernanda Costa Brandão Berti, que estiveram muito presentes nestes anos finais de doutorado, sempre me apoiando e incentivando.

Agradeço muito aos meus eternos amigos Heloísa Magagnin Brincas e Guilherme de Souza Nogueira por todos esses anos de amizade, companheirismo, por sempre me apoiarem durante os anos de graduação, mestrado e doutorado. Vocês são parte dessa conquista.

À minha avó, Maria do Carmo Pizani Marrara, pela amizade, pelo amor e por sempre cuidar de mim. À minha tia, Maria Cecília Marrara, por ter sido mais que uma tia ao longo

desses anos, pela amizade, pelo companheirismo, pelas risadas e por tudo o que representa na minha vida.

À minha querida Suelen Cristina Soares Baal por ter entrado de forma tão essencial na minha vida e por sempre me incentivar nesta reta final.

Em especial, aos meus pais Silvia Helena Marrara Mathias e José Luiz Mathias por tudo o que fizeram por mim. Por sempre me incentivarem a seguir meu sonho. Por todo o carinho e amor que têm por mim. Sem o apoio de vocês essa conquista não seria possível.

Finalmente, agradeço a todos que de alguma forma fizeram parte do desenvolvimento desse trabalho, em especial as pacientes que cederam o material para o nosso laboratório.

RESUMO

O câncer de mama é o tipo de neoplasia mais frequente entre as mulheres no mundo e compreende um grupo heterogêneo de doenças classificadas em diferentes subtipos. A classificação comumente utilizada na prática clínica é baseada em marcadores imunoistoquímicos. Nessa classificação os tumores de mama são divididos em quatro grupos principais: luminal A, luminal B, HER2 positivo e triplo negativo. Com o aprimoramento das tecnologias de larga escala, desenvolveram-se classificações baseadas no perfil de expressão de RNAs codificadores. Entretanto, sabe-se que 98% do genoma humano não codifica para proteínas, evidenciando a relevância de se investigar RNAs não codificantes (ncRNAs). Os ncRNAs são classificados de acordo com o seu tamanho, e os RNAs longos não codificantes (lncRNAs) compreendem uma classe de transcritos que apresentam mais de 200 nucleotídeos. Estas moléculas apresentam alta versatilidade, atuando na regulação da expressão gênica em diversos níveis. Além disso, estas moléculas apresentam alta especificidade temporal e local, possibilitando que sejam utilizadas como biomarcadores. Os lncRNAs já foram descritos como desregulados em diversos tipos de câncer, incluindo o de mama. Porém, pouco se sabe ainda do papel funcional destes lncRNAs nos diferentes subtipos da doença. Desta forma, o presente trabalho teve como objetivo identificar lncRNAs reguladores de redes transcricionais no câncer de mama utilizando metodologias de bioinformática e de validação funcional. Por meio de uma metodologia de predição de redes de co-expressão, identificou-se a rede mediada pelo LINC00504 como diferencialmente ativada entre os subtipos luminal A e basal-*like*. O silenciamento deste lncRNAs em linhagens celulares do subtipo luminal A resultou na diminuição da proliferação celular e da capacidade clonogênica das células, sugerindo o papel oncogênico do LINC00504 nestas linhagens. Empregou-se também uma metodologia de análise bioinformática para evidenciar lncRNAs relevantes na resposta imune tumoral nos diferentes subtipos moleculares. O método foi bastante eficiente, e foi possível evidenciar lncRNAs específicos para cada subtipo: LINC01871 associado com ativação da resposta imune e melhor prognóstico no subtipo basal-like, EBLN3P que foi relacionado com supressão da resposta imune no subtipo luminal B, e MEG3, XXYLT1-AS2, e LINC02613 foram associados com a ativação da resposta imune nos subtipos luminal A, HER2-*enriched* e normal-*like* respectivamente. A partir dos resultados obtidos neste trabalho, pode-se enfatizar a relevância de melhor caracterizar os lncRNAs nos diferentes subtipos do câncer de mama, fornecendo assim novas perspectivas em relação ao diagnóstico, prognóstico e alvos terapêuticos.

Palavras-chave: Câncer de mama, lncRNAs, Bioinformática, Biomarcadores

ABSTRACT

Breast cancer is the most common type of cancer among women worldwide and comprises a heterogeneous group of diseases classified into different subtypes. The classification commonly used in clinical practice is based on immunohistochemical markers. In this classification, breast tumors are divided into four main groups: luminal A, luminal B, HER2 positive and triple-negative. With the improvement of sequencing technologies, classifications based on the expression profile of coding RNAs were developed. However, it is known that 98% of the human genome does not code for proteins, evidencing the relevance of investigating non-coding RNAs (ncRNAs). ncRNAs are classified according to their size, and long non-coding RNAs (lncRNAs) comprise a class of transcripts that have more than 200 nucleotides. These molecules have high versatility, acting in the regulation of gene expression at different levels. Furthermore, these molecules have high temporal and local specificity, allowing them to be used as biomarkers. lncRNAs have been described as dysregulated in several types of cancer, including breast cancer. However, little is known about the functional role of these lncRNAs in the different subtypes of the disease. Thus, this study aimed to identify lncRNAs that regulate transcriptional networks in breast cancer using bioinformatics and functional validation methodologies. Through a method of prediction of co-expression networks, the network mediated by LIN00504 was identified as differentially activated between luminal A and basal-like subtypes. The silencing of these lncRNAs in luminal subtype A cell lines resulted in a decrease in cell proliferation and clonogenic capacity of the cells, suggesting the oncogenic role of LINC00504 in these cell lines. A bioinformatics analysis methodology was also used to reveal relevant lncRNAs in the tumor immune response in different subtypes. The method was very efficient, and it was possible to show specific lncRNAs for each subtype: LINC01871 associated with activation of the immune response and better prognosis in the basal-like subtype, EBLN3P, which was related to the suppression of the immune response in the luminal B subtype, and MEG3, XXYLT1 -AS2, and LINC02613 were associated with immune response activation in luminal A, HER2-enriched and normal-like subtypes respectively. From the results obtained in this work, the need and relevance of better characterizing lncRNAs in different breast cancer subtypes can be emphasized to provide new perspectives regarding diagnosis, prognosis and therapeutic targets.

Keywords: Breast cancer, lncRNAs, Bioinformatics, Biomarkes.

LISTA DE ILUSTRAÇÕES

REVISÃO DE LITERATURA

CAPÍTULO II

CAPÍTULO III

LISTA DE TABELAS

REVISÃO DE LITERATURA

CAPÍTULO II

LISTA DE SIGLAS

°C – grau Celsius

µg – micrograma

µL – microlitro

µm – micrômetro

BC – Breast cancer

Cat# - Catalog number

DEPC - Diethyl pyrocarbonate

DNA - Deoxyribonucleic Acid

dNTPs - Deoxynucleotide Triphosphate

FBS – Fetal Bovine Serum

g – g force

h – hours

HER2 – Human Epidermal Growth Factor Receptor 2

KI-67 - Proliferation-related ki-67 antigen

lncRNA – Long non-coding RNA

LumA – Luminal A

LumB – Luminal B

min - Minutos

miRNA - microRNAs

mL – mililitro

nM – nanomolar

pmol – picomol

RE – Receptor de estrógeno

RNA - Ribonucleic Acid

RNA-Seq – RNASequencing

RP – Receptor de progesterona

RT-qPCR - Reverse Transcriptase quantitative Polymerase Chain Reaction

SUMÁRIO

1. INTRODUÇÃO

O câncer de mama é o tipo de neoplasia mais incidente no mundo, representando 11,7% de todos os casos de câncer. Entre as mulheres, o câncer de mama é responsável por 1 em cada 4 casos de câncer e 1 em cada 6 mortes por câncer, ocupando o primeiro lugar em incidência na grande maioria dos países. (SUNG *et al.,* 2021). De acordo com o Instituto Nacional do Câncer (INCA, 2020), foram esperados para o ano de 2020 o diagnóstico de 66.280 casos de câncer de mama no Brasil. O crescente número de casos diagnosticados desta neoplasia no Brasil e no mundo ao longo dos anos caracteriza o câncer de mama como um problema de saúde pública, que exige a superação de barreiras sociais, econômicas e psicológicas. Ampliar os conhecimentos sobre sua etiologia, desenvolver alternativas de tratamento e novos métodos de diagnóstico precoce são prioridades na oncologia.

Levando em conta a heterogeneidade do câncer de mama, são utilizadas na prática clínica ferramentas que possibilitam maior precisão no diagnóstico da doença. Parâmetros clínicos (classificação TNM e grau histológico) e de biomarcadores imunoistoquímicos [receptor de estrogênio (RE), receptor de progesterona (RP), e receptor do fator de crescimento epidermal (*HER2*)] são amplamente utilizados e facilitam a estratificação das pacientes em subgrupos. Entretanto, estas classificações não são suficientes para prever o prognóstico e indicar o melhor tratamento para todas as pacientes.

A classificação de câncer de mama utilizando-se de critérios moleculares de expressão gênica de mRNAs foi estabelecida inicialmente por Perou *et al*., (2000) e por Sorlie *et al*, (2003). Nestes primeiros trabalhos, os tumores de mama foram subdivididos em cinco grupos principais: Luminal A, Luminal B, HER2-*enriched*, Basal-*like* e Normal-*like.*

A utilização da classificação molecular possibilitou o desenvolvimento de tratamentos específicos para cada subtipo da doença, contribuindo para a sobrevida geral e para melhor qualidade de vida das pacientes. Entretanto, a classificação molecular baseada apenas em dados de expressão de mRNAs não é suficiente para o acompanhamento de todas as pacientes, sendo ainda necessária a busca por novos marcadores moleculares.

A classificação molecular foi essencialmente desenvolvida utilizando-se dados de expressão de sequências codificadoras de proteínas, que compreendem apenas 2% do genoma humano (LANDER *et al*., 2001; VENTER *et al*., 2001). Já foram estimados que 98% dos transcritos do genoma humano são RNAs não codificantes (ncRNAs) (DIAMANTOPOULOS *et al.,* 2018). Classificações moleculares do câncer de mama que utilizam estas moléculas de RNA estão em crescente investigação.

Entre as moléculas de ncRNAs, destacam-se os RNAs longos não codificantes, que compreendem um subtipo de ncRNAs com mais de 200 nucleotídeos, e que estão envolvidos em diversos processos biológicos fundamentais para o desenvolvimento do câncer (YANG *et al.,* 2013; XING *et al.,* 2014)

O papel regulatório dos lncRNAs ainda é pouco conhecido. Sabe-se que estes RNAs podem atuar em diferentes níveis de regulação da expressão gênica. Este potencial regulatório deve-se à versatilidade da molécula, que tem capacidade de se dobrar em estruturas secundárias, ligando-se em diversos substratos de forma altamente específica (GEISLER; COLLER, 2013). Além disso, os lncRNAs exibem um padrão de expressão tecido-específico, o que evidencia sua utilização como marcadores de diagnóstico e de classificação molecular (WU; DU, 2017).

Sendo assim, o presente trabalho teve como objetivo identificar lncRNAs reguladores que atuam em processos biológicos fundamentais no desenvolvimento do câncer de mama. Para isto, foram empregadas metodologias de bioinformática na busca por redes de co-expressão mediadas pelos lncRNAs e por lncRNAs relacionados com a resposta imune nos subtipos moleculares. Os resultados obtidos neste trabalho estão apresentados na forma de artigos científicos nos capítulos I, II e III.

2. REVISÃO DE LITERATURA

2.1 Câncer de Mama

O câncer de mama representa 25% de todos os casos de câncer diagnosticados no mundo (FERLAY *et al.,* 2018). No Brasil, o Instituto Nacional do Câncer (INCA) estimou para o ano de 2020 o diagnóstico de 66.280 casos da doença.

O desenvolvimento do câncer de mama pode ocorrer em qualquer tipo celular da mama, e representa uma doença bastante heterogênea, exibindo diferentes características morfológicas, perfis imunoistoquímicos, diferentes subtipos histopatológicos e moleculares.

A classificação do carcinoma mamário tem bastante importância para determinação de prognóstico e tratamento das pacientes. Primeiramente, é realizada a identificação do sítio de origem da proliferação celular desordenada. Existem dois principais sítios de origem da doença: o revestimento de ductos mamários, que caracteriza os carcinomas ductais, e dos lóbulos mamários, os carcinomas lobulares (MAKKI, 2015).

Além da identificação do sítio de origem da neoplasia, também é determinada a capacidade migratória das células, caracterizando carcinomas *in situ*, em que a proliferação de células epiteliais limitada pela extensão dos ductos ou lóbulos,ou invasivo, que apresenta capacidade de invadir a membrana basal e colonizar tecidos distantes do local de origem do tumor primário (BATEMAN; SHAW, 2016).

A necessidade de se classificar os casos de câncer em grupos apareceu com a observação de que as taxas de sobrevida eram diferentes quando se dividia as pacientes de acordo com algumas características. O sistema de classificação TNM, se baseia na extensão anatômica da doença, considerando as características do tumor primário (T), as características dos linfonodos das cadeias de drenagem em que o tumor se localiza (N) e a presença ou ausência de metástases à distância (M). A esses parâmetros são adicionados números indicando a extensão da doença (DENOIX,1952).

Os tumores de mama são também classificados de acordo com o grau de diferenciação histológica, com base em parâmetros de diferenciação celular,

formação do túbulo/glândula, pleomorfismo nuclear e contagens mitóticas, baseado no sistema de classificação de Bloom e Richardson (1957) modificado por Elston e Ellis (1991). Neste modelo de classificação, tumores classificados com grau I (baixo) caracterizam-se por serem bem diferenciados e apresentarem melhor prognóstico, grau II (intermediário) moderadamente diferenciado, e grau III (alto), pouco diferenciado e pior prognóstico.

Os primeiros trabalhos que analisaram o perfil de expressão gênica e diversidade fenotípica do câncer de mama foram realizados por Perou e colaboradores (2000), e posteriormente validado por Sorlie e colaboradores em anos seguintes (2001 e 2003). Nesta classificação, o câncer de mama é subdividido em cinco subtipos moleculares: luminal A, luminal B, HER2 *enriched*, basal-*like,* e normal-*like*. Mais recentemente dois novos subtipos foram identificados como: claudina baixa e molecular apócrino (SANGA *et al.,* 2009; PRAT *et al.,* 2010).

Atualmente, com base na análise de assinatura molecular em microarranjos que utiliza como base um conjunto de genes mínimo (PAM50) e com maior aplicação clínica, pacientes diagnosticadas com câncer de mama são classificadas em cinco subtipos principais: luminal A, luminal B, basal-*like,* HER2-*enriched* e normal-*like.* Estes subtipos apresentam características biológicas e epidemiológicas específicas, que podem ajudar a explicar as diferenças no comportamento dos tumores, e resposta ao tratamento, apesar da aparente semelhança morfológica, auxiliando no tratamento personalizado do paciente (NIELSEN *et al.,* 2010; HENGE *et al.,* 2017).

O subtipo luminal A representa 50-60% dos casos de câncer de mama. Apresenta expressão de receptor estrogênio e progesterona positivos, e baixos níveis de expressão de Ki67. Representa o melhor prognóstico entre os subtipos de câncer de mama, e seu índice de recidiva é de aproximadamente 5%, sendo significativamente menor que os outros subtipos (EROLES *et al.,* 2012).

Comparativamente com o subtipo luminal A, o luminal B apresenta um fenótipo mais agressivo, pior prognóstico, maior grau histológico, maior índice proliferativo (Ki67 >14%) representando 10-20% dos casos. Quanto ao seu perfil imunoistoquímico, este subtipo apresenta expressão de receptor de estrogênio e expressão variável do receptor de progesterona e HER2. Aproximadamente 8% dos

casos diagnosticados como luminal B apresentam recorrência da doença (EROLES *et al.,* 2012).

O subtipo HER2-enriched corresponde a cerca de 15-20% dos casos diagnosticados. É bem definido pela alta expressão do gene *HER2* e dos genes que estão associados à sua via. Em relação às características morfológicas, tumores do subtipo HER2-*enriched* apresentam altos índices de proliferação, 75% exibem alto grau histológico, exibindo pior prognóstico. Estima-se que 13% destes casos apresentam recidiva da doença.

Tumores classificados como basal-*like* representam aproximadamente 10-20% dos casos de câncer de mama, sendo mais comuns em mulheres mais jovens. Apresentam alto grau histológico, desenvolvem frequentemente metástase em linfonodos axilares e apresentam um prognóstico pior quando comparado com os subtipos luminais e cerca de 17% dos casos exibem recidiva (EROLES *et al.,* 2012).

O subtipo normal-*like* corresponde a aproximadamente 5-10% dos casos. E apresenta perfil de expressão gênica semelhante ao da mama normal. No que se refere ao prognóstico, este subtipo tem um pior prognóstico comparado ao subtipo luminal A (DAI *et al.,* 2015).

A partir do aprimoramento de tecnologias de alto rendimento para análise de expressão gênica, passou-se a empregar o conceito de que o câncer de mama representa um conjunto de diferentes doenças que afetam o mesmo órgão e que provém de um mesmo sítio de origem (o ducto, por exemplo), mas apresentam diferentes características histopatológicas, de prognóstico e de resposta ao tratamento (BRITT *et al.,* 2020). A classificação molecular é um aspecto de extrema importância para o prognóstico das pacientes, entretanto, a implementação desta classificação nos sistemas de saúde não é algo viável a curto e médio prazo.

Dessa forma, foi sugerida uma classificação que leva em consideração características clínico-patológicas, que são estratificadas de acordo com a análise imunoistoquímica dos tumores, uma vez que este método é mais viável para a implementação nos sistemas de saúde pública. Por meio dessa análise é possível realizar uma aproximação dos subtipos moleculares, de acordo com a expressão proteica dos seguintes marcadores: receptor de estrógeno e progesterona , HER2 e o marcador de proliferação celular Ki-67. Nestes modelos de classificação os tumores são divididos em quatro subtipos principais: luminal A, luminal B, HER2 positivo e triplo negativo. São considerados tumores de melhor prognóstico os subtipos luminais A e B e de pior prognóstico HER2 positivo e triplo negativo (GOLDHIRSCH *et al.,* 2013).

A classificação imunoistoquímica de acordo com os parâmetros utilizados, bem como informações de prognóstico, está resumida no quadro 1.

Subtipo	ER/PR/HER2	Ki-67	Prognóstico
Luminal A	ER+ PR+ HER2-	<14%	Bom
Luminal B	ER+ PR+ HER2+	<14%	Intermediário
	ER+ PR+ HER2-	>14%	Desfavorável
HER2 Positivo	ER- PR- HER2+	>14%	Desfavorável
Triplo Negativo	ER- PR- HER2-	>14%	Desfavorável

FONTE: Adaptado de Goldhirsch *et al.,* 2013

O manejo atual de pacientes diagnosticadas com câncer de mama ainda depende da avaliação patológica tradicional. Pouco ainda é empregado na clínica no que se refere ao diagnóstico utilizando-se de biomarcadores em nível genômico. Ainda se faz necessário maior esforço para tornar esses avanços tecnicamente convenientes e disponíveis para uso clínico, adicionalmente a identificação de novos marcadores que possam facilitar a ampla utilização (BRITT *et al.,* 2020).

2.2 RNAS LONGOS NÃO CODIFICANTES

i. 2.2.1 Origem e Biogênese

Por muito tempo a ciência esteve focada no estudo de sequências de DNA que eram transcritas em RNA e subsequentemente traduzidas em proteínas específicas, porém, sabe-se atualmente que aproximadamente 2% do genoma humano é codificante (MATICK, 2001). Adicionalmente, o aprimoramento das técnicas de sequenciamento em larga escala possibilitou a identificação de inúmeros novos transcritos do genoma humano, com mais de 98% sendo de RNAs não codificantes (DIAMANTOPOULOS *et al.,* 2018)

Os RNAs não codificantes são divididos em dois principais grupos: (1) ncRNAs curtos que apresentam menos de 200 nucleotídeos (nt), compreendendo os microRNAs, importantes componentes da regulação pós-transcricional da expressão gênica, os piwiRNAs que estão relacionados com a regulação de atividade de transposons, e também os snRNAs que desempenham papel de regulação do spliceossomo; (2) ncRNAs longos (lncRNAs), que apresentam mais de 200 nt e estão envolvidos na regulação da expressão gênica em variados níveis (CLARK;MATICK, 2011).

A origem e evolução dos lncRNAs é discutida de forma bastante ampla e alguns cenários são hipotetizados neste contexto (Figura 1). O primeiro destes cenários é a ocorrência de mutação disruptiva em genes codificadores de proteínas (Figura 1A). Como exemplo deste tipo de mecanismo, pode-se citar o lncRNA XIST, essencial para o cromossomo X em mamíferos eutérios. Recentemente, descobriu-se que o XIST e sua sequência promotora foram derivados de um gene codificador de proteínas denominado *Lnx3* que adquiriu mutações em seu quadro de leitura durante o período de evolução dos mamíferos placentários (DURET *et al.,* 2006, ELISAPHENKO *et al.,* 2008). LncRNAs podem também ser originados como resultado de rearranjos cromossômicos (Figura 1B). Neste caso, duas regiões genômicas previamente separadas se tornam justapostas e dão origem a um lncRNA (PONTING *et al.,* 2009). Eventos de duplicação gênica (Figura 1C) e inserção de elementos transponíveis (Figura 1D) são também mecanismos propostos relacionados com origem e evolução dos lncRNAs (PONTING *et al.,* 2009).

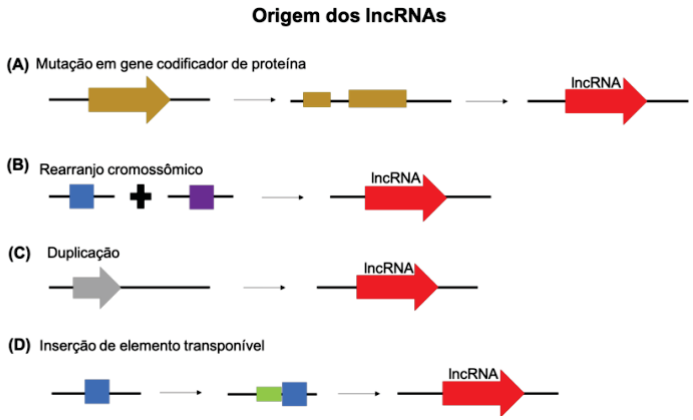


FIGURA 1 – REPRESENTAÇÃO DAS HIPÓTESE EVOLUTIVAS RELACIONADAS COM A ORIGEM DOS lncRNAs; (A) lncRNAs podem se originar através de mutações em genes codificadores de proteínas; (B) Regiões previamente separadas se tornam justapostas após rearranjos cromossômicos originando lncRNAs; (C) Duplicações gênicas em tandem podem também originar lncRNAs; (D) Inserção de elementos transponíveis origina um lncRNA funcional. FONTE: A Autora, 2021. Adaptado de Ponting et al., 2009.

Detalhando características globais de lncRNAs e mRNAs sugere-se que os genes de lncRNA são menos conservados evolutivamente, em termos de sequência de nucleotídeo, contém menos exons, são menos abundantemente expressos e apresentam alta especificidade celular e temporal (GUO *et al.,* 2020). Semelhante aos mRNAs, os lncRNAs são transcritos pela RNA polimerase II, apresentam CAP-5' e cauda poli-A 3' (BUNCH, 2018). No entanto, estudos recentes começaram a revelar processos distintos de transcrição, processamento, exportação dos lncRNAs, que estão intimamente ligados a seus destinos e funções celulares.

Semelhante ao observado em mRNAs, os lncRNAs apresentam enriquecimento de H3K4me3 em regiões promotoras; no entanto genes de lncRNAs tem maiores níveis de H3K27ac e são mais fortemente reprimidos por complexos de remodelamento de cromatina como Swr1, Isw2 e Rsc (QUINN; CHANG,2016). A baixa expressão dos lncRNA deve-se também a uma menor sobreposição de motivos de ligação de fatores de transcrição em regiões promotoras de lncRNAs (MATTIOLI *et al.,* 2019).

Eventos de *splicing* também são observados em moléculas de lncRNAs, porém de forma menos eficiente do que é observado para mRNAs. As sequências de lncRNAs têm sinais de *splicing* mais fracos e distâncias maiores entre sítios 3' e os pontos de quebra (GUO *et al.,* 2020).

No citoplasma, lncRNAs são expressivamente menos abundantes que mRNAs. Estima-se que 4% dos lncRNAs estão enriquecidos no citoplasma, enquanto 26% corresponde a mRNAs (DERRIEN *et al.,* 2012). A retenção no núcleo

dos lncRNAs deve-se a diversos fatores: *splicing* ineficiente, presença de sítios de poliadenilação e presença de sequências internas que promovem a retenção nuclear. Por exemplo, o lncRNA MEG3 contém uma sequência de 356 nucleotídeos que promove a retenção nuclear ao se associar com U1 snRNP (AZAM *et al.,* 2019).

2.2.2 Classificação e Função

A classificação dos lncRNAs ainda é bastante discutida no meio científico sendo que as formas mais utilizadas de classificação envolvem a posição genômica do lncRNA em relação ao gene codificante mais próximo e seu papel funcional.

De acordo com a classificação baseada em sua posição genômica, lncRNAs podem ser organizados em seis grupos principais: senso, antisenso, intergênico, intrônico, bidirecional e acentuador (MA *et al.,* 2013) (Figura 2).

LncRNAs classificados como senso são transcritos na mesma orientação que o gene codificador de proteínas mais próximo podendo estar contidos em uma parte deste gene ou estar completamente sobreposto (Figura 2A). Ao contrário, lncRNAs classificados como antisenso são transcritos em orientação oposta ao gene codificador mais próximo (Figura 2B). Estima-se que aproximadamente 32% dos lncRNAs humanos são classificados como antisenso de gene codificadores (DERRIEN *et al.,* 2012). Os lncRNAs classificados como intergênicos, também denominados lincRNAs, são regulados transcricionalmente de forma similar aos mRNAs e se encontram entre genes codificadores (MA *et al.,* 2013) (Figura 2C). Os intrônicos são transcritos a partir dos íntrons de genes codificadores (Figura 2D); os bidirecionais estão localizados na posição oposta dos genes que codificam proteínas e são transcritos da proximidade (menos de 1000 pares de bases) de seus genes vizinhos (Figura 2E) e por fim, os acentuadores são transcritos a partir das regiões acentuadoras (*enhancers*) de genes codificantes (Figura 2F) (KIM *et al.,* 2015).

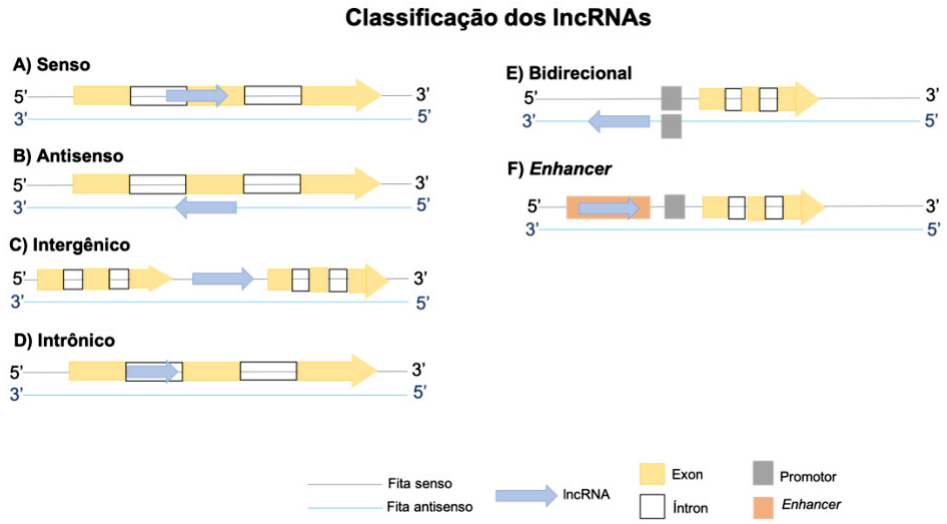


FIGURA 2 - CLASSIFICAÇÃO DOS lncRNAS DE ACORDO COM SUA POSIÇÃO GENÔMICA. A) lncRNAs senso são transcritos na mesma orientação do gene codificador de proteínas mais próximo; B) ao contrário os antisenso são transcritos na orientação oposta;C) os lncRNAs intergênicos estão dentro de regiões intergênicas; D) os intrônicos são transcritos a partir de íntrons do gene codificante; D) os bidirecionais são localizados na região oposta e transcritos próximo de seus genes vizinhos; E) os lncRNAs enhancers são transcritos a partir de regiões potenciadoras (enhancers) de genes codificadores de proteínas. FONTE: A Autora, 2021. Adaptado de Liu et al., 2019.

Os lncRNAs podem também ser classificados de acordo com seu mecanismo molecular de ação, conforme ilustrado na Figura 3. LncRNAs podem atuar como sinais moleculares (Figura 3), sendo expressos em tempo e espaços específicos; refletindo as ações combinatórias de fatores de transcrição ou vias de sinalização para indicar a regulação gênica no espaço e no tempo. Como exemplos de atuação como sinais, pode-se citar o lncRNAs PANDA e lincRNA-p21 que são induzidos pela proteína p53 em resposta ao dano no DNA (HUNG *et al.,* 2011, HUARTE *et al.,* 2010). A função de *decoy*, ou isca, refere-se à capacidade de lncRNAs sequestrarem fatores de transcrição, modificadores de cromatina ou outros fatores regulatórios para longe do seu local de ação, desta forma, reprimindo de forma indireta a transcrição (WANG; CHANG, 2011). Como exemplo de lncRNA atuando como isca, pode-se citar o "*telomeric repeat-containing RNA*" (TERRA). O lncRNA TERRA é intimamente relacionado com a proteção e regulação dos telômeros. Este lncRNA se associa fisicamente a transcriptase reversa da telomerase a subunidade catalítica do complexo da telomerase que adiciona sequências de repetição telomérica para proteger as extremidades dos cromossomos do desgaste (FLYNN *et al.,* 2011).

Ao atuarem como guias, os lncRNAs ligam-se às proteínas e, em seguida, direcionam a localização do complexo de ribonucleoproteína para alvos específicos, podendo atuam em *cis* ou em *trans*, podendo ativar ou reprimir a transcrição, dependendo se os complexos guiados são fatores de transcrição ou complexos repressivos (WANG; CHANG, 2011). O HOTAIR é um exemplo de lncRNA que atua como guia. Este lncRNA

se liga ao complexo repressivo *polycomb* (PRC2), atuando no silenciamento do locus *HOXD*, quando este se encontra super-expesso (BALAS *et al.,* 2018).

Tradicionalmente, as proteínas eram consideradas as principais participantes em vários complexos organizados em *scaffolding* (arcabouço) (GOOD *et al.,* 2011). Evidências recentes, no entanto, revelam que os lncRNAs desempenham papel semelhante (Figura 3). Moléculas efetoras e repressoras se ligam em domínios específicos dos lncRNAs formando complexos ribonucleoproteicos, podendo atuar na regulação da cromatina, de estruturas nucleares ou complexos de sinalização (WANG; CHANG, 2011). O lncRNA telomérico TERC é um exemplo clássico de um arcabouço ao montar o complexo da telomerase que mantém as extremidades dos telômeros, combinando a atividade da transcriptase reversa com proteínas de alvos do telômero em um complexo ribonucleoprotéicos (BALAS *et al.,* 2018)

LncRNAs *enhancers* são moléculas de ação *cis* que são geradas a partir de elementos potenciadores (*enhancers*) ativos e necessários para a função do potenciador e subsequente ativação da expressão gênica de genes vizinhos que codificam proteínas (ØROM *et al.,* 2010). Como exemplo deste tipo de mecanismo, pode-se citar a ativação do proto-oncogene *SPHK1,* que em resposta a proliferação celular forma uma hélice tripla a montante do potenciador *SPHK1*, que ajuda a recrutar modificadores da cromatina que ativam a transcrição de eRNA-*SPHK1* e promovem a expressão de *SPHK1* (Statello *et al.,* 2020)

As principais funções dos lncRNAs desempenhadas no citoplasma estão representadas na Figura 3. Os lncRNAs podem ter como alvo, moléculas de mRNAs modulando sua estabilidade, atuando no decréscimo e aumento da expressão da molécula de mRNA. Como exemplos, pode-se citar os lncRNAs 1/2-sbsRNAs e BACE1-AS. Os lncRNAs 1/2-sbsRNAs promovem a ligação de proteínas STAU1, conhecidas como *RNA-binding proteins* envolvidas no processo de degradação de mRNAs (RASHID *et al.,* 2016). O lncRNA BACE1-AS promove o aumento da expressão do mRNA BACE1 mascarando os sítios de ligação ao miR-485-5p (FAGHIHI *et al.,* 2010). Como arcabouços no citoplasma, os lncRNAs se ligam a proteínas impedindo que as mesmas se liguem a molécula alvo (RINN; CHANG, 2012), como discutido previamente no núcleo.

Outra função bastante relevante desempenhada pelos lncRNAs no citoplasma refere-se à competição endógena. ranscritos de RNA codificantes e não codificantes competem pela ligação a miRNAs. LncRNAs podem apresentar diversos sítios de ligação para miRNAs, e o sequestro destes miRNAs resulta no aumento da expressão de moléculas de mRNAs alvos (SALMENA *et al.,* 2011).

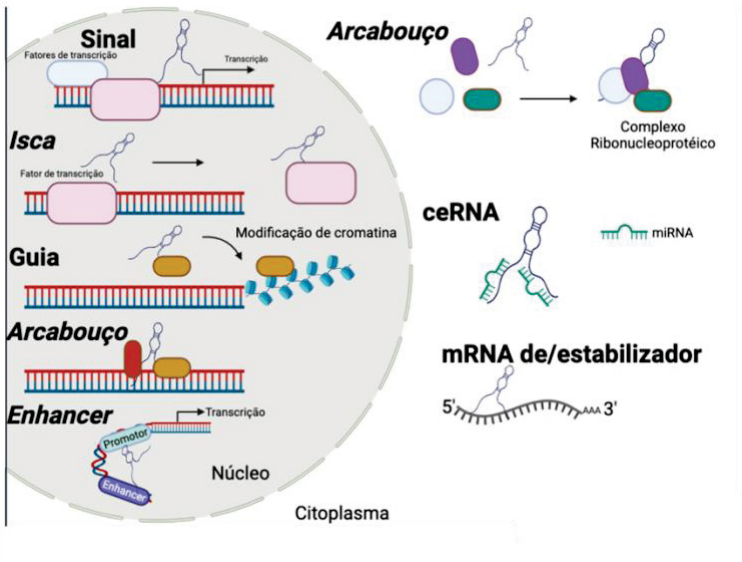


FIGURA 3 – REPRESENTAÇÃO DE MECANISMOS MOLECULARES MEDIADOS PELOS lncRNAs. No núcleo, os lncRNAs podem atuar como sinais, guias, decoy, scaffold ou enhancers. Estes mecanismos relacionam-se com a regulação transcricional da célula.No citoplasma, estão evidenciadas as funções de arcabouço, cerna e de/estabilizador de mRNAs. FONTE: A Autora,2021. Adaptado de Bar et al., 2016.

2.2.3 LncRNAs e a Regulação da Expressão Gênica

LncRNAs exibem uma alta versatilidade ao regular a expressão gênica em diversos níveis: modelamento da cromatina, transcricional, pós-transcricional, traducional e pós-traducional.

Modificações no estado da cromatina e das histonas afetam diretamente a acessibilidade das regiões gênicas. Diversos lncRNAs estão localizados na cromatina e podem atuar por dois mecanismos centrais: interagindo com proteínas, facilitando ou inibindo sua ligação e atividade em regiões de DNA alvo (SALDAÑA-MEYER *et al.,* 2019); e interagindo com a molécula de DNA, formando uma molécula híbrida, em forma de triplexes ou *R-loops*, promovendo ou inibindo a acessibilidade da cromatina (KUO *et al.,* 2019).

A regulação transcricional mediada por lncRNAs pode ser desempenhada por alguns mecanismos principais: O lncRNA pode recrutar um complexo de proteína regulatória em *cis* ou *trans,* inibindo a ligação de proteínas regulatórias atuando como "*decoy*", organizando regiões hetero ou eucromáticas alterando a conformação, e o ato de transcrição ou *splicing* do lncRNA pode gerar um estado de cromatina ou impedimento estérico, influenciando a expressão de genes próximos (LONG *et al.,* 2017; GIL;ULITSKY, 2020).

No que se refere à regulação pós-transcricional, por meio de pareamento de bases, os lncRNAs podem estabilizar ou promover a tradução de mRNAs alvo, enquanto o pareamento parcial facilita o decaimento do mRNA ou inibe a tradução do mRNA alvo. Na ausência de complementaridade, os lncRNAs podem suprimir o *splicing* e a tradução do mRNA precursor, agindo como iscas de proteínas de ligação ao RNA ou miRNAs.

Um dos mecanismos mais amplamente estudados de regulação pós-transcricional, refere-se a competição endógena em que lncRNAs e mRNAs competem pela ligação de miRNAs alterando a expressão do mRNA alvo (YOON *et al.,* 2013; SALMENA *et al.,* 2011).

A regulação da tradução mediada por lncRNAs ocorrer em diversos pontos da tradução por meio da regulação da expressão e função de fatores traducionais, podendo atuar tanto na promoção como na inibição deste processo (KARAKAS; OZPOLAT, 2021). Por fim, observa-se que lncRNAs estão envolvidos em várias modificações pós-traducionais de proteínas, principalmente fosforilação, ubiquitinação e acetilação, regulando assim a degradação ou formação de proteínas e afetando seus níveis de expressão (ZHANG *et al.,* 2019).

2.2.4 Papel dos lncRNAs nos "*Hallmarks*" do Câncer

Por muito tempo, os lncRNAs foram negligenciados por serem considerados ruídos transcricionais. No entanto, o papel regulatório destes transcritos passou a chamar bastante atenção em vários processos biológicos, além de estarem associados com doenças multifatoriais (CIPOLLA *et al.,* 2018), incluindo o câncer (DE OLIVEIRA *et al.,* 2019).

A desregulação de lncRNAs vem sendo observada em vários tipos de câncer, e associada a vários processos celulares, como transcrição, tráfego intracelular e remodelamento cromossômico. As marcas registradas ("*Hallmarks*") do câncer, descritas por Hanahan e Weinberg (2000), compreendem capacidades celulares adquiridas que são comumente observadas no processo de carcinogênese. Estão incluídos, a sustentação da sinalização proliferativa, evasão de supressores de crescimento, resistência à morte celular, imortalidade replicativa, indução da angiogênese e ativação de invasão / metástase. Além disso, em 2011, instabilidade do genoma, inflamação, reprogramação do metabolismo energético e mecanismos de evasão do sistema imune foram adicionados como importantes aspectos durante o desenvolvimento de várias etapas de tumores humanos (HANAHAN; WEINBERG,2011).

Os lncRNAs são evidenciados em todas as características adquiridas do câncer, conforme representado na Figura 4. Neste trabalho, destaca-se o "*Hallmark*" relacionado com evasão do sistema imune, estando os resultados de lncRNAs envolvidos neste processo, no câncer de mama, descritos no capítulo III.

O microambiente imunológico tumoral (MIT) desempenha um papel importante na progressão do câncer de mama, e devido à heterogeneidade dos subtipos da doença, são observados distintos perfis imunológicos, que podem afetar diretamente o prognóstico das pacientes. As células reguladoras presentes no MIT, como células T reguladoras (Treg), macrófagos associados a tumor (TAM), células supressoras mielóide (MDSC) e as células T auxiliares do tipo 2 (Th2) estão associadas a um microambiente imunológico supressor e piores prognósticos, enquanto outras células, incluindo linfócitos T citotóxicos (CTL), células T auxiliares tipo 1 (Th1) e células assassinas naturais (NK) estão associados a um microambiente imunológico antitumoral e resultados favoráveis (TAMIMI *et al.,* 2008; DEL ALCAZAR *et al.,* 2017).

Infiltrados de células regulatórias incluindo Tregs, MDSC, Th2, Th17, M2 macrófagos, células T HLADR− e células Tγδ estão relacionadas com fenótipo mais agressivo, especialmente em RE− e subtipo triplo-negativo, sugerindo o microambiente

mais imunossupressor nesses casos (SADEGHALVAD *et al.,* 2021). Em contraste, células NK, CTL, Th1, e as células Tfh estão relacionadas à atividade antitumoral, sendo observadas mais frequentemente em subtipos RE+. A presença destas células em subtipos RE- foi também associada com melhor prognóstico (SADEGHALVAD *et al.,* 2021).

Utilizando dados de imunogenômica, Thorsson e colaboradores (2018) caracterizaram o microambiente tumoral de vários tipos de câncer, incluindo o câncer de mama. A partir de uma análise integrada de dados, os autores classificaram os tumores em seis subtipos imunes principais, que eles chamaram de C1-C6. Esses subtipos têm conjuntos distintos de assinaturas imunológicas, que também podem estar relacionados ao prognóstico. C1 (cicatrização de feridas) exibiu elevada expressão de genes angiogênicos, uma alta taxa de proliferação e um viés de células Th2 para o infiltrado imune adaptativo. C2 (IFN-γ dominante) apresentou um sinal T CD8 + forte, maior diversidade TCR e uma alta taxa de proliferação. C3 (inflamatório) foi o subtipo que apresentou genes Th17 e Th1 altos e baixos para proliferação moderada de células tumorais. C4 (depleção de linfócitos) exibiu uma assinatura de macrófago mais proeminente com Th1 resposta M2 suprimida e alta. C5 (imunologicamente silencioso) foi enriquecido por tumores cerebrais e exibiu o menor linfócito e a maioria das respostas aumentadas de macrófagos. Finalmente, C6 (TGF-β dominante) exibiu a maior assinatura de TGF-β e um alto infiltrado linfocítico com uma distribuição uniforme de células T tipo I e tipo II (THORSSON *et al.,* 2018)

De acordo com esta abordagem, o câncer de mama foi classificado em cinco subtipos (C1, C2, C3, C4 e C6), sendo C2 (n = 345) o subtipo mais representativo, seguido por C1 (n = 320). Os subtipos imunes também foram descritos de acordo com os subtipos moleculares, e como esperado, os subtipos variaram significativamente de acordo com esses grupos imunes. Por exemplo, o subtipo luminal A foi mais representativo do subtipo C1, enquanto as amostras basais de C2 (THORSSON *et al.,* 2018).

Neste trabalho, entretanto, os autores focaram suas análises em mRNAs nos subtipos imunes propostos. Sabe-se, porém, que os ncRNAs, em especial os lncRNAs, desempenham funções regulatórias no que se refere ao sistema imune. Pode-se citar como exemplos, o lncRNA HOTAIR que foi identificado induzindo a diferenciação de granulócitos (ZHANG *et al.,* 2014) e o lncRNA CD244 que regula TNF-α através da modulação da metilação da cromatina do gene *PCR2*; inibindo a expressão de IFN-γ (KANG *et al.,* 2017).

A importância funcional dos lncRNAs relacionados ao sistema imunológico está apenas começando a ser caracterizada; seu papel como biomarcador e alvo na imunoterapia do câncer precisa ser examinado mais detalhadamente, e neste sentido, foi explorado o perfil de lncRNAs envolvidos nos subtipos imunes relacionados ao câncer de mama, conforme descrito no capítulo III deste trabalho.

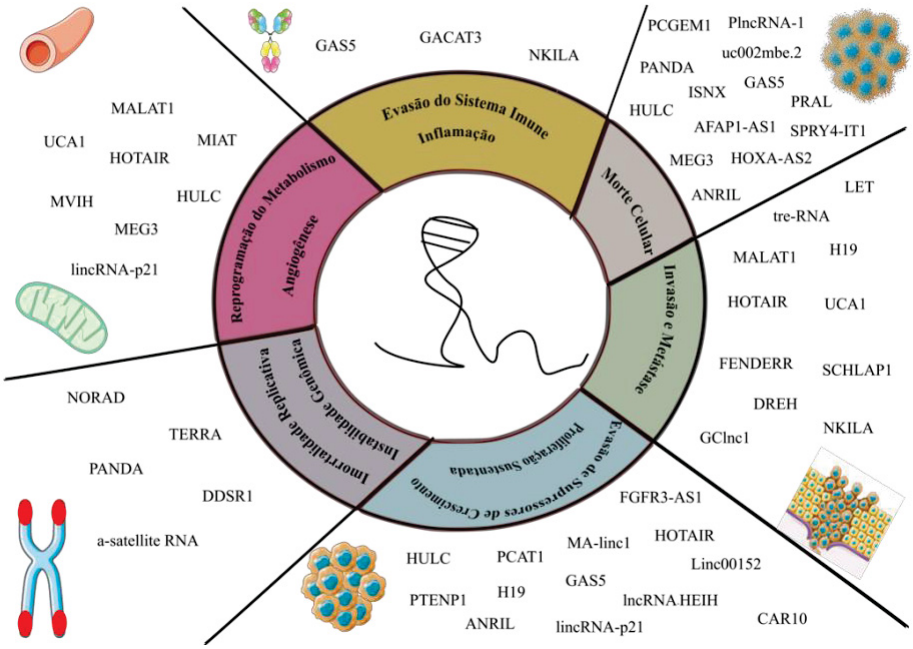


FIGURA 4- REPRESENTAÇÃO DE lncRNAS ENVOLVIDOS NOS “HALLMARKS” DO CÂNCER. Diversos lncRNAs são conhecidamente associados com os “hallmarks” do câncer. Alguns deles, como por exemplo, HOTAIR foi evidenciado em mais de um processo celular. FONTE: Adaptado de Garcia *et al.,* 2019

3. OBJETIVOS

Identificar RNAs longos não codificantes (lncRNAs) que atuem como reguladores de redes transcricionais diferencialmente ativas em subtipos de câncer de mama.

3.1 OBJETIVOS ESPECĪFICOS

- Computar possíveis redes de co-expressão que sejam mediadas por lncRNAs no câncer de mama a partir de dados armazenados no banco de dados The Cancer Genome Atlas (TCGA);
- Identificar redes de co-expressão que apresentem diferença de atividade (ativa ou inativa) nos diferentes subtipos de câncer de mama;
- Avaliar o efeito da modulação (silenciamento/hiperexpressão) *in vitro* de lncRNAs selecionados como potenciais reguladores em diferentes linhagens de câncer de mama na regulação de processos celulares importantes como proliferação, apoptose e ciclo celular;
- Identificar lncRNAs que atuem na resposta imune tumoral;
- Investigar o papel dos lncRNAs relacionados com a resposta imune nos diferentes subtipos moleculares do câncer de mama.

4. MATERIAL E MÉTODOS

A metodologia descrita a seguir (4.1 – 4.10) refere-se aos resultados apresentados no capítulo II.

4.1 Análise de Redes de Co-Expressão

A matriz de expressão de 12 mil lncRNAs da coorte de pacientes com câncer de mama do TCGA foi extraída do banco de dados TANRIC. Entre o total de lncRNAs foram selecionados inicialmente em que a expressão fosse detectada em pelo menos 90% das amostras, resultando em 3.680 lncRNAs.

Para construção das redes de co-expressão foi utilizada a pipeline presente no pacote RTN (CASTRO *et al.,* 2016). Nesta metodologia, as redes de co-expressão são construídas a partir de uma métrica chamada informação mútua (I.M) acompanhada de análise de permutação e *bootstrap*. As redes de co-expressão dos 3.680 lncRNAs foram computadas utilizando-se três níveis de significância: $p<10^{-6}$, $p<10^{-7}$ e $p<10^{-8}$. A fim de garantir um maior nível de estringência da análise, foram utilizadas as redes regulatórias computadas no nível de significância de $p<10^{-8}$.

Após a computação das redes de co-expressão dos 3.680 lncRNAs, foram retirados da análise redes com tamanho inferior de 15 elementos e também aquelas com *Differential Enrichment Score* (dES) -0,5 < dES > 0,5, evitando assim, a seleção de redes com pequeno valor informativo. Esta seleção resultou em um número final de 84 redes que também foram estratificados de acordo com o subtipo de câncer: basal-*like*, HER2-*enriched*, luminal A e luminal B. Amostras normal-*like* foram retiradas da análise devido ao pequeno número amostral (n=7).

Para seleção das redes de co-expressão, foi realizado o teste não paramétrico de Kruskal-Wallis e posterior teste de Dunn para verificar a diferença da variância de dES comparando os subtipos de câncer de mama para as 84 redes selecionadas. Neste passo, filtramos as diferenças significativas tomando como base o subtipo basal-like e valor de $p<10^{-4}$ resultando em 40 redes de co-expressão. Estas redes foram então avaliadas quanto valor médio de dES em cada subtipo de câncer de mama. Foram considerados valores informativos -0,5≤ dES ≥0,5 e que apresentassem a maior diferença de valor entre os subtipos luminal A e basal-like, por serem os subtipos com maiores variações de prognóstico.

4.2 Caracterização da Amostra

Foram utilizadas 55 amostras de pacientes com câncer de mama, predominantemente europeias (SUGITA *et al.,* 2016), sendo 30 classificadas como luminal A e 25 como triplo negativo. negativo provenientes do hospital Nossa Senhora da Graças, Curitiba, Paraná, Brasil. As amostras de tecido foram obtidas de procedimento cirúrgico armazenadas em RNA *later* e -80ºC. Todas as amostras foram obtidas com consentimento das pacientes através da assinatura do termo de consentimento livre e esclarecido.

As amostras foram classificadas a partir de dados de imunoistoquímica de acordo com Goldhirsch et al. 2013. Este projeto foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos do Setor de Ciências da Saúde, UFPR, sob o número de CAAE:19870319.3.0000.0102.

4.3 Isolamento de RNA e Síntese de cDNA

O isolamento de RNA das amostras de tecido e das linhagens celulares foi realizado utilizando-se o *RecoverAll Total Nucleic Acid Isolation Kit* (Invitrogen) e TRIzol® (Invitrogen), respectivamente, seguindo as recomendações do fabricante.

O total de 1 µg do RNA extraído foram submetidos a tratamento com DNAse I (Invitrogen). Para cada amostra foi adicionado 1µl DNAse buffer 10x (500nM TRIS, ph 8, 100nM MgCl2), 1µl de DNAse I (1 U) e 5µl de água Milli-Q tratada com dietilpirocarbonato (DEPC) e incubado por 30 min a 37°C. Ao final, adicionou-se 1µl de EDTA (50mM) e incubado a 65°C por 10min.

As amostras tratadas com DNAse I foram então convertidas em cDNA utilizando-se o kit *Super Script III Reverse Transcriptase* (Invitrogen). Para cada amostra, primeiramente foi adicionado 1µL de dNTP (10mM), 1µL de *Random Primers*, 12µL de RNA, numa concentração de cerca de 1µg. Essa reação foi então levada ao termociclador por 5 min a 65ºC e por 2 min a 4ºC. O passo posterior envolve a transcrição reversa, em que foi adicionado 4µL do Buffer 5x, 1µL de DTT (0,1M), 0,5µL da enzima Super Script III (200U/µL) e 0,5 de RNase out. (5000U). Este mix foi incubado em termociclador por 5 min a 25ºC, 60 min a 50ºC e 15 min a 70ºC.

4.4 PCR Quantitativa em Tempo Real

As reações de PCR foram realizadas utilizando-se o master *mix Power SYBR Green PCR Master Mix* (Applied Biosystems) e os primers apropriados para cada reação (listados no material suplementar do capítulo II). Os experimentos foram realizados em triplicata utilizando o aparelho Viia-7 (Applied Biosystems). Para cada reação foi realizado controle RT-.

A quantificação relativa foi avaliada a partir da comparação do ciclo em que o nível de fluorescência alcançou um limiar pré-estabelecido (ct) do gene alvo em relação a dois controles endógenos, β-glucuronidase (*GUS-β*) e *β*-actina (*ACTB*).

Para os cálculos de expressão nas amostras de tecido de pacientes de câncer de mama um *pool* de cDNA de linhagens celulares de câncer de mama foi utilizado como amostra calibradora. No experimento de silenciamento do lncRNA os mesmos controles endógenos foram utilizados, e como amostra calibradora o cDNA da linhagem celular transfectada com o siRNA controle negativo. A quantificação do gene alvo, normalizado em relação ao gene endógeno e em relação ao calibrador foi obtida pela fórmula $2^{-\Delta\Delta CT}$.

4.5 Cultivo celular

Para os experimentos utilizando cultivo celular foram utilizadas linhagens luminal A MCF-7 e ZR-75-1. As células foram cultivadas em meio RPMI (Gibco) suplementado com 10% de SFB (Gibco) e 1% Penicillin-Streptomycin. Todas as células foram mantidas em uma estufa úmida com 5% de CO_2 a 37°C. Após atingir confluência necessária de cerca de 80%, as células foram tripsinizadas para utilização em experimentos posteriores.

4.6 Transfecção do siRNA

Um total de $6x10^5$ MCF-7 e ZR75-1 foram plaqueadas em placas de seis poços até atingirem confluência total de 50-70%. As células foram transfectadas com 125 ηM de siRNA controle negative (CN) ou siRNA LINC00504 (ThermoFisher – número 4392420). Para transfecção foi utilizado o protocolo de transfecção reversa, em que as células são colocadas em suspensão com o siRNA, utilizando-se o reagente Lipofectamina 2000 (Invitrogen). As células foram avaliadas após 24h, 48h e 72h.

4.7 Viabilidade Celular

Para o ensaio de viabilidade celular, as células MCF-7 e ZR75-1 transfectadas com siRNA CN e siRNA LINC00504 foram plaqueadas em placas de 96 poços. Para cada poço foram adicionadas 500 células. A viabilidade celular foi avaliada após 48, 72 e 96 horas de transfecção. Em cada dia, o meio de cultura RPMI foi substituído por 100μl de RPMI + 10μl de corante Resazurina (7-hidroxi-3H-fenoxazin-3-ona 10-óxido) na concentração de 0,1 mg / ml e incubados por 4 horas. A leitura foi realizada no sistema TECAN Infinite® 200 PRO em 600 ηm e 570 ηm. O experimento foi feito em triplicata biológica e as leituras em quadruplicata técnica.

4.8 Capacidade Clonogênica

As células MCF-7 e ZR75-1 transfectadas foram plaqueadas em placas de seis poços (1000 células / poço), e cresceram por 25 dias a 37 ° C em uma incubadora umidificada com 5% de CO_2. Após este período, as células foram fixadas com metanol 100% por 20 min, coradas com cristal violeta 1% por 5 min e lavadas com água até o excesso de corante ser removido (todas as etapas foram realizadas à temperatura ambiente) e posteriormente contou-se o número de colônias em cada poço. O experimento foi realizado em triplicata biológica.

4.9 Apoptose

A taxa de apoptose foi analisada usando o kit de detecção de apoptose Annexin V-FITC (Sigma-Aldrich). Quarenta e oito horas após a transfecção, as células MCF-7 e ZR75-1 foram ressuspendidas em tampão de ligação contendo Anexina V-FITC e P iodeto de propídio, de acordo com instruções do fabricante. As amostras foram analisadas por citometria de fluxo (B.D. Biosciences, EUA), e foram discriminadas em células viáveis, células necróticas, apoptose precoce e células de apoptose posterior usando o citofluorímetro BD FACSVantage ™ (B.D. Biosciences, EUA). Em seguida, as porcentagens de células apoptóticas de cada grupo foram comparadas. A análise dos dados foi realizada usando o software FlowJo (v.10). O experimento foi realizado em triplicata biológica.

4.10 Análises Estatísticas

Os dados são apresentados como média ± D.P. A análise estatística dos dados foi realizada pelo teste t de Student usando R (versão 3.6.3) (http: ///www.r-project.org/). Valores de $p \leq 0.05$ foram considerados estatisticamente significativos.

A metodologia descrita a seguir (4.11 – 4.14) refere-se aos resultados apresentados no capítulo III.

Análise de Assinaturas de lncRNAs

4.11 Dados Imunes do Câncer de Mama

As informações dos subtipos moleculares, imunes, dados de sobrevida e da fração leucocitária do câncer de mama foram extraídas do material suplementar de Thorsson e colaboradores (2018). De acordo com o código da amostra, foi possível integrar dados de expressão gênica e de subtipo imunológico. Os subtipos moleculares do câncer de mama foram definidos de acordo com a classificação PAM50. Para esta análise, foram considerados subtipos imunes que apresentassem mais de cinco pacientes. Grupos com menos de cinco pacientes foram excluídos da análise.

4.12 Obtenção de dados de RNA-Seq

Os dados de expressão gênica de pacientes com câncer de mama foram extraídos do banco de dados TCGA. A matriz de expressão gênica total foi obtida por meio do repositório *XenaBrowser* no link https://xenabrowser.net/datapages/. Os dados foram extraídos normalizados na versão log2 FPKM ("*Fragments Per Kilobase Million*") e integrados com os dados de subtipos imunes.

A anotação dos lncRNAs na matriz de expressão gênica foi feita utilizando o pacote *biomaRt* v2.46.3 (DUNRICK et al., 2005).

4.13 Seleção dos lncRNAs Específicos de Subtipos Moleculares e Imunes

Os lncRNAs foram selecionados a fim de obter moléculas específicas em cada subtipo molecular que permitisse a diferenciação entre os subtipos imunes. A seleção dos lncRNAs foi realizada utilizando-se uma métrica denominada "*Signal-to-noise ratio*" (SNR). Nesta métrica, a média de expressão do lncRNA no grupo analisado é subtraída da média de expressão nos outros grupos e dividida pela soma dos desvios padrão dos grupos, conforme a fórmula abaixo:

$$SNR = \frac{\mu 1 - \mu 2}{\sigma 1 + \sigma 2}$$

A utilização desta métrica, permite a seleção de lncRNAs contrastantes entre os grupos excluindo-se o ruído da técnica de sequenciamento.

SNR foi aplicado em dois momentos na análise: primeiramente, aplicou-se a métrica para selecionar lncRNAs específicos de cada subtipo molecular; e

posteriormente para selecionar lncRNAs específicos de cada subtipo imune. Nestas seleções, aplicou-se filtros quantílicos: entre os subtipos moleculares foram selecionados lncRNAs que se encontravam acima de quantil 95% e nos subtipos imunes os que estavam acima do quantil 98%. Foi utilizada a plataforma *InteractVenn* ((http://www.interactivenn.net/) para selecionar lncRNAs específicos em cada etapa.

Os lncRNAs selecionados no final desta análise de duas etapas foram representados graficamente utilizando-se o pacote *ComplexHeatmap* (GU *et al.,* 2016).

4.14 Correlação com a Fração Leucocitária

Os valores de expressão dos lncRNAs selecionados foram utilizados para calcular a correlação com a fração leucocitária dentro de cada grupo. Utilizou-se o método de Spearman e o valor de p foi ajustado com o método *False Discovery Rate* (FDR).

4.15 Análise de Sobrevida

Os lncRNAs selecionados foram avaliados em relação à sobrevida dos pacientes nos subtipos moleculares. Para isto, foi aplicado o método de regressão de Cox baseado na expressão dos lncRNAs para avaliar sobrevida livre de doença e intervalo de progressão da doença. Utilizou-se o pacote *survival v3.2-10* e os valores de p ajustados de acordo com o método FDR. Valores de FDR abaixo de 10% foram considerados significativos.

Dentro de cada subtipo molecular foi selecionado um lncRNA baseado nos valores de p da regressão de Cox e da fração leucocitária para ser também avaliado com o método de Kaplan-Meier. Utilizou-se o valor da mediana da expressão do lncRNA para dividir os pacientes em dois grupos (acima e abaixo da mediana) e o valor de p foi calculado pelo teste de *log rank* utilizando o pacote *survival*.

4.16 Análise de Enriquecimento de Vias

Análises de enriquecimento de vias foram realizadas em três etapas: a) primeiramente dividiu-se os pacientes, dentro de cada subtipo molecular, em dois grupos de acordo com o valor de expressão do lncRNA selecionado tendo o valor da sua mediana como base; b) aplicou-se o método do SNR para os dados de expressão dos mRNAs nos grupos determinados e c) os valores de SNR dos mRNAs foram ordenados e conduzida a análise de *Gene Set Enrichment Analysis* (GSEA) utilizando-se o pacote *fgsea* v1.16.0 (KOROTKEVICH et al., 2016) e as assinaturas gênicas denominadas “Hallmarks” do banco de dados MSigDb v7.2 (Subramanian et al., 2005; Liberzon et al., 2015).

5. DESCRIÇÃO DOS CAPÍTULOS

A parte de Resultados e Discussão desta tese será dividida em três capítulos organizados na forma de artigos científicos:

O capítulo I apresenta o artigo de revisão ***"Long non-coding RNAs differential expression in breast cancer subtypes: What do we know?"*** publicado em 2018 na revista *Clinical Genetics.* O artigo foi anexado neste documento conforme as normas da revista. Este artigo apresenta uma revisão bibliográfica sobre lncRNAs nos diferentes subtipos do câncer de mama. Nesta revisão, ressaltamos a grande heterogeneidade apresentada entre os subtipos da doença, uma vez que os lncRNAs evidenciados foram, na maioria das vezes, específicos para cada subtipo da doença.

O capítulo II, apresenta o artigo original: ***"Novel lncRNAs Co- Expression Networks Identifies LINC00504 with Oncogenic Role in Luminal A Breast Cancer Cells"*** publicado em 2021 na revista *International Journal of Molecular Sciences*. O artigo foi anexado neste documento conforme as normas da revista. Este artigo descreve a metodologia bioinformática empregada neste trabalho na busca de lncRNAs candidatos em diferentes subtipos moleculares do câncer de mama, e na sequência apresentamos os resultados da validação funcional, em cultura celular, de um lncRNA alvo selecionado. Este lncRNA, denominado LINC00504, apresentou potencial papel oncogênico em linhagens luminal A de câncer de mama, já que seu silenciamento reduziu viabilidade celular e capacidade clonogênica das células.

O capítulo III apresenta o artigo original "***Unraveling immune related lncRNAs in breast cancer molecular subtypes"*** publicado em 2021 na revista *Frontiers in Oncology.* Neste artigo, foram investigados lncRNAs relacionados com diferentes respostas imunes nos subtipos moleculares do câncer de mama.

6. CAPÍTULO I

Arquivo completo não disponível devido à regra de direitos autorais da revista.

Received: 6 October 2018 | Revised: 28 December 2018 | Accepted: 31 December 2018

DOI: 10.1111/cge.13502

WILEY CLINICAL GENETICS

REVI EW

Long non-coding RNAs differential expression in breast cancer subtypes: What do we know?

Carolina Mathias[1] | Erika P. Zambalde[1] | Philip Rask[2] | Daniela F. Gradia[1] | Jaqueline C. de Oliveira[1]

[1]Department of Genetics, Federal University of Parana, Curitiba, Brazil

[2]Department of Experimental Therapeutics, University of Texas MD Anderson Cancer Center, Houston, Texas

Correspondence
Department of Genetics, Federal University of Parana, 81530-000 Curitiba, Brazil.
Email: carol.mathias1@hotmail.com

Breast Cancer (BC) is the most commonly diagnosed cancer and is the leading cause of cancer deaths in women. BC is a heterogeneous disease with different clinical and genetic features. According to immunohistochemical markers, BC is subdivided into four main subtypes: luminal A, luminal B, ERBB2 positive and triple negative. Long non-coding RNAs (lncRNAs) are transcripts with more than 200 nucleotides and deregulated lncRNAs are associated with human diseases, including BC. In order to improve BC molecular classification, non-coding RNAs (ncRNAs), including lncRNAs, have been used. In this review, we focus on lncRNAs with differ- ential expression in BC subtypes and how these RNAs may act to contribute to BC heterogene- ity. We also emphasize the potential of these lncRNAs as biomarkers.

KEY WORD S

breast cancer, breast cancer subtypes, ER positive, lncRNA

end of the article

How to cite this article: Mathias C, Zambalde EP, Rask P, Gradia DF, de Oliveira JC. Long non-coding RNAs differential expression in breast cancer subtypes: What do we know? *Clin Genet*. 2019;1-11. https://doi.org/10.1111/cge.13502

7. CAPÍTULO II

International Journal of
Molecular Sciences

MDPI

Article

Novel lncRNAs Co-Expression Networks Identifies LINC00504 with Oncogenic Role in Luminal A Breast Cancer Cells

Carolina Mathias [1], Clarice S. Groeneveld [2,3], Sheyla Trefflich [4], Erika P. Zambalde [1], Rubens S. Lima [5], Cícero A. Urban [5], Karin B. Prado [1], Enilze M. S. F. Ribeiro [1], Mauro A. A. Castro [4], Daniela F. Gradia [1] and Jaqueline C. de Oliveira [1,*]

1 Post-Graduation Program in Genetics, Department of Genetics, Federal University of Parana, Curitiba 81530-900, PR, Brazil; carol.mathias1@hotmail.com (C.M.); erikazambaldi@gmail.com (E.P.Z.); kbraun@ufpr.br (K.B.P.); enilzeribeiro@gmail.com (E.M.S.F.R.); danielagradia@gmail.com (D.F.G.)
2 Cartes d'Identité des Tumeurs Program, Ligue Nationale Contre le Cancer, 75013 Paris, France; clari.groeneveld@gmail.com
3 Oncologie Moleculaire, Institut Curie, CNRS, UMR144, Equipe Labellisée Ligue Contre le Cancer, 75005 Paris, France
4 Bioinformatics and Systems Biology Laboratory, Polytechnic Center, Federal University of Parana (UFPR), Curitiba 81520-260, PR, Brazil; sheylatrefflich@gmail.com (S.T.); mauro.a.castro@gmail.com (M.A.A.C.)
5 Breast Disease Center, Hospital Nossa Senhora das Graças, Curitiba 80810040, PR, Brazil; rsilveiralima@uol.com.br (R.S.L.); cicerourban@hotmail.com (C.A.U.)
* Correspondence: jaqueline.carvalho@ufpr.br

check ror updates

Citation: Mathias, C.; Groeneveld, C.S.; Trefflich, S.; Zambalde, E.P.; Lima, R.S.; Urban, C.A.; Prado, K.B.; Ribeiro, E.M.S.F.; Castro, M.A.A.; Gradia, D.F.; et al. Novel lncRNAs Co-Expression Networks Identifies LINC00504 with Oncogenic Role in Luminal A Breast Cancer Cells. *Int. J. Mol. Sci.* **2021**, *22*, 2420. https://doi.org/10.3390/ijms22052420

Received: 29 December 2020
Accepted: 25 January 2021
Published: 28 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affil- iations.

Abstract: Long non-coding RNAs (lncRNAs) are functional transcripts with more than 200 nucleotides. These molecules exhibit great regulatory capacity and may act at different levels of gene expression regulation. Despite this regulatory versatility, the biology of these molecules is still poorly understood. Computational approaches are being increasingly used to elucidate biological mechanisms in which these lncRNAs may be involved. Co-expression networks can serve as great allies in elucidating the possible regulatory contexts in which these molecules are involved. Herein, we propose the use of the pipeline deposited in the RTN package to build lncRNAs co-expression networks using TCGA breast cancer (BC) cohort data. Worldwide, BC is the most common cancer in women and has great molecular heterogeneity. We identified an enriched co-expression network for the validation of relevant cell processes in the context of BC, including LINC00504. This lncRNA has increased expression in luminal subtype A samples, and is associated with prognosis in basal-like subtype. Silencing this lncRNA in luminal A cell lines resulted in decreased cell viability and colony formation. These results highlight the relevance of the proposed method for the identification of lncRNAs in specific biological contexts.

Keywords: LINC00504; breast cancer; co-expression; lncRNA; luminal A

1. Introduction

Non-coding RNAs are a big class of transcripts that can be classified according to their size, comprising small RNAs <200 nucleotides (nt) and long non-coding RNAs (lncRNA)
>200 nt [1]. LncRNA molecules are usually transcribed by RNA polymerase II, capped, and polyadenylated with some being also spliced. LncRNAs present high tissue specificity and great regulatory versatility, acting at different levels of gene expression regulation [2,3]. LncRNAs have already been analyzed in several human diseases, including cancer, with varying regulatory activity as either oncogenic or tumor suppressor, whose

activity can modulate all hallmarks of cancer [4]. For example, the lncRNA HOTAIR can promote tumor growth and metastasis in several cancer types, such as breast, hepatocellular, lung and gastric cancer. [5].One lncRNA known to act as a tumor suppressor, regulating p53, is the Maternally expressed gene 3 (MEG3). Several studies have shown down-regulation of this lncRNA in human cancers, such as lung, breast, gastric and colorectal [6]

LncRNAs display a highly tissue-specific expression, which can render them potential candidates for cancer diagnosis. Their expression also correlates with overall survival (OS), metastasis, tumor stage or grade, highlighting theirpotential use as prognostic markers [7]. Nevertheless, only a small portion of lncRNAs have been functionally characterized, and the function of most of them remains elusive. To better understand these molecules' biology, various computational methods and tools for identifying, annotating, and performing functional prediction for long non-coding RNAs are being used extensively [8].

One of the most used approaches towards understanding the function of lncRNAs is based on co-expression patterns shared with protein-coding counterparts [9,10]. Varied methodologies have already been used to predict these co-expression networks; however, caution should be exercised when analyzing large patient cohorts' data.

In this study, we propose to use the pipeline available from the Bioconductor/R package RTN. This method is tuned to deal with large gene expression datasets to build transcriptional regulatory units. It was previously used to compute networks regulated by transcriptional factors [11,12] and uses mutual information (MI) metric based on a gene's expression varying across a cohort. In the present work, we used this approach to construct regulatory networks focused on units of co-expression networks due to the fact of its great statistical rigor for its analysis.

Breast cancer (BC) is a critical health problem worldwide. GLOBOCAN 2018 estimated the diagnosis of 2,088,849 new cases worldwide [13]. BC is a heterogeneous disease characterized by several pathological features, divergent treatment responses, and substantial differences in long-term patient survival [14]. According to gene expression profiles, BC may be classified mainly into four molecular subtypes, luminal A and B, HER2-enriched and basal-like. Luminal A and B are both sensitive to hormone therapy, although lu- minal A patients have often low-grade tumors and good prognosis while luminal B is recognized as having a higher proliferation rate. HER2-enriched and basal-like types are widely considered to have poorer survival and tumors with higher grade [15]. In relation to therapeutic conduct, positive estrogen subtypes benefit from treatment by Tamoxifen and HER2-enriched patients can benefit from monoclonal antibody therapy. Patients with basal-like tumors have poorer survival [16,17], and remain the most challenging group to treat, but new targeted therapies are becoming available, as, for example, PARP inhibitors in *BRCA* mutated patients. Here, we used lncRNA co-expression networks to identify important lncRNAs in BC by focusing on differences between molecular subtypes.

The results obtained through the use of the proposed methodology highlight the potential biological relevance of lncRNAs in BC. Therefore, this methodology could be used for screening of functional lncRNAs in different physiological and pathological contexts.

2. Results

2.1. LncRNAs Co-Expression Networks Activity Are Different among Breast Cancer Molecular Subtypes

To build the co-expression networks, we used data from 12,000 lncRNAs identified in the TCGA cohort. Since lncRNAs intrinsically have low expression, we selected only lncR- NAs expressed in at least 90% of the tumor samples. This filter resulted in 3680 lncRNAs, that were then used to build the co-expression networks, in addition to mRNA expression data. In this way, co-expression networks could jointly present lncRNAs and mRNAs in their composition.

Next, we looked for networks that had characteristics intrinsic to their statistical construction, which could reflect greater biological relevance in the analyzed context. The networks were first filtered according to their size, those containing less than 15 genes were exceeded, and then, according to Differential Enrichment Score (dES) $0.5 < dES > 0.5$ values, obtaining 84 lncRNA co-expression networks (Supplementary Table S2). As represented in Figure 1, many differences can be identified, mainly between the luminal A and basal-like subtypes, which make up the most clinically contrasting subtypes. The most significant

contrasts between these subtypes are emphasised in the figure. To identify lncRNAs co-expression networks significantly different between these two subtypes, we used KruskalWallis tests to statistically determine the co-expression networks differentially expressed in these two subtypes using the dES metrics. At this point, the data were filtered according to the higher numeric differences between the luminal A and basal-like subtypes due to the great clinical contrast of these subtypes. In this last step, we obtained a list of 40 lncRNA co-expression networks (Supplementary Table S3).

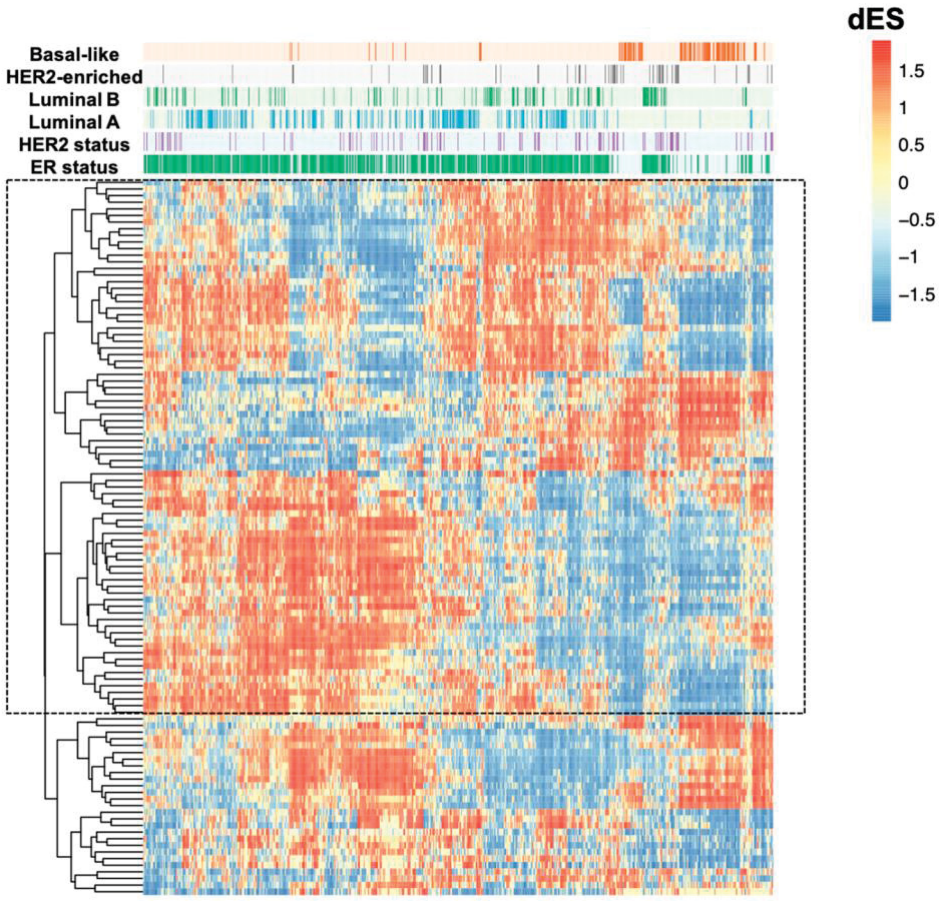


Figure 1. Heatmap representation of the 84 lncRNA co-expression networks activity in breast cancer molecular subtypes. According to this metric, positive dES (in red) are activated; and negative dES (in blue) are inactivated. In the highlighted rectangle are the more contrasting between luminal A and basal-like subtypes. The samples are organized in the heatmap columns and categorized according to the annotation bar at the top of the graph. The TCGA BRCA cohort samples were classified according to their molecular classification and estrogen receptor and HER2 status. LncRNAs in their co-expression networks (n = 84) are clustered in the rows and can be found in Supplementary Table S2.

2.2. LncRNAs Co-Expression Networks Are Enriched for Cancer-Related Function

The 40 co-expression networks filtered, according to criteria described in the previ- ous results section, were analyzed for biological process enrichment. This enrichment analysis was performed using the molecular signature data "Hallmarks" deposited in the MSigDB database. Among these co-expression networks, 21 of them were statistically enriched (p < 0.01) for some cancer-related function (Figure 2).

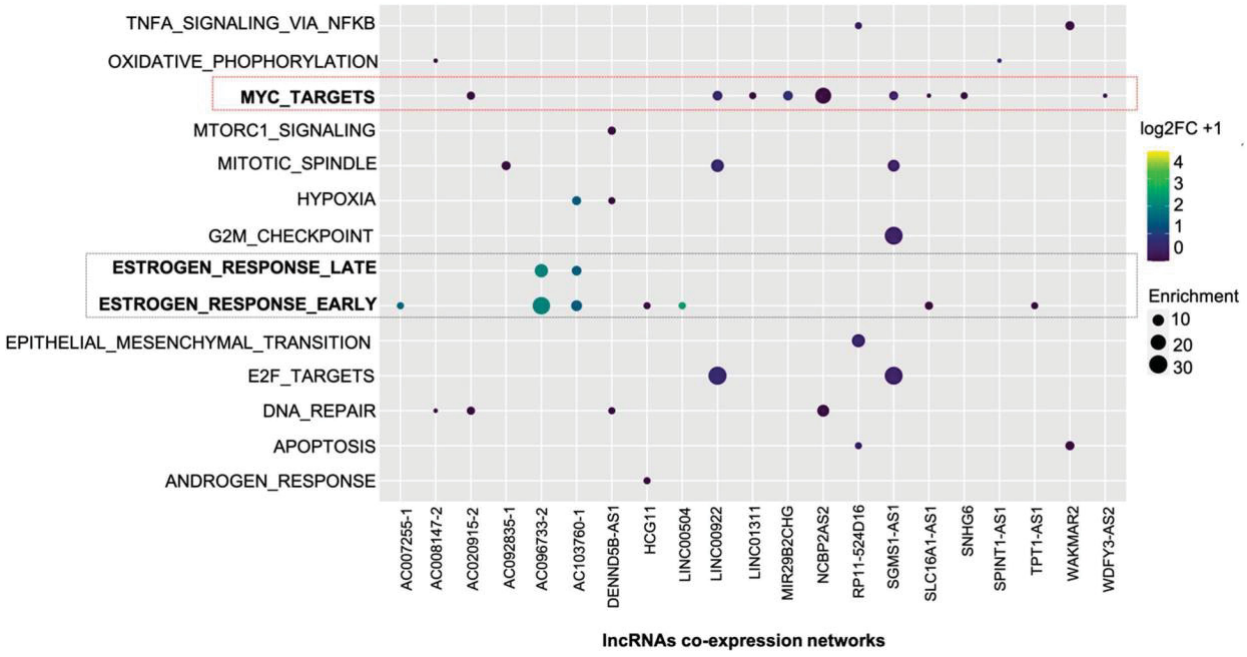


Figure 2. Enrichment map of the identified co-expression networks. In rows are the 14 identified molecular signatures, and in the columns, the networks of co-expression of 21 lncRNAs. Dot size refers to the number of genes identified in each cancer-related process. Blank spaces mean that there was no statistically significant enrichment. Dot color refers to the difference (logFC + 1) of the lncRNAs used as a reference for building co-expression networks. The comparison was made between luminal A and basal-like, with dots with colors closer to yellow more expressed in luminal subtype A. Colors closer to purple indicate down-regulation in luminal A. In this enrichment map, the pathway named "MYC_TARGETS" (red rectangle) and "ESTROGEN_RESPONSE_EARLY" (black rectangle) stands out as the most frequently identified.

Among the 50 datasets available in "Hallmarks" collection, 14 were identified in the analyzed co-expression networks. The most frequently identified were "MYC_TARGETS" (n = 9) and ESTROGEN_RESPONSE_EARLY (n = 7). The results obtained highlight the consistency of the RTN pipeline's data since the estrogen and *MYC* oncogene signaling pathways are known to be relevant in the process of mammary carcinogenesis.

To biologically validate the relevance of one of the lncRNAs used in constructing the co-expression networks, we chose the one that presented the wider dES range comparing the luminal A and basal-like subtypes. This lncRNA is named LINC00504 (ENSG00000248360), and according to Figure 2, it has a significative value for "ESTROGEN_RESPONSE_EARLY" enrichment.

2.3. LINC00504 Co-Expression Network in Breast Cancer Luminal-A and Basal-Like Subtypes

We organized LINC00504 co-expression network data in a ranked dES plot for the BC cohort (Figure 3A). This graph highlights the activation of these networks in luminal A patients and positive for estrogen and progesterone receptors. Figure 3A also shows that basal-like patients are grouped in negative network activity (represented in blue). Genes already positively related to luminal subtype A, such as transcription factors (TFs) *ESR1*, *GATA3*, and *FOXA1*, can be recognized. According to our data, these TFs have positive mutual information values with LINC00504, meaning the concordant expression variation among these molecules (Figure 3B). The graphical representation of the co-expression network was limited to genes with high mutual information values for better visualization.

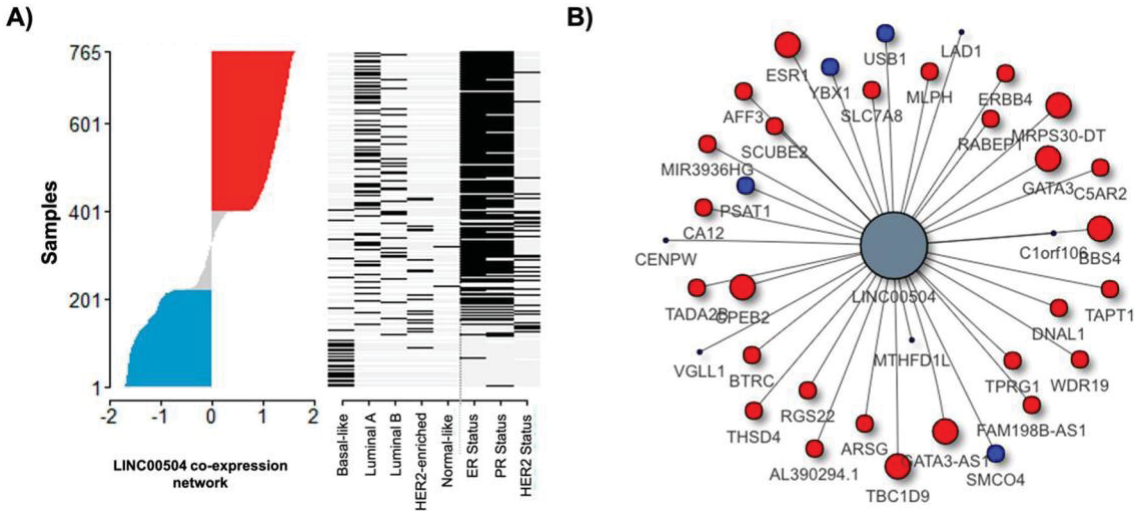


Figure 3. Network activity and LINC00504 co-expression network representation. (**A**) Ranked dES plot for the BC cohort and status of key attributes plot. Each dash in the status attributes plot represents a patient. The cohort of patients was extracted according to their molecular classification and immunohistochemical parameters (estrogen and progesterone receptors and HER2 status). Positive dES, indicated by the color red, represents the up regulation of this network along the samples, whereas negative dES is indicated by the color blue. In this case, positive regulon activity is located in Luminal A patients; (**B**) Network representation of LINC00504 zoomed in on genes with higher mutual information values. Red circles refer to genes with positive correlation and blue to negative ones. Dot size refers to the mutual information value, where big dots indicates higher mutual information.

2.4. LINC00504 Is Up-Regulated in Luminal A and Is Related to Better Prognosis in Basal-Like Subtype

After identifying LINC00504 as a potential relevant lncRNA in mammary carcinogenesis, we focused on the functional characterization of the LINC00504. Since its co-expression network has been identified with the significant difference between the luminal A and basal-like subtypes, we first evaluated its expression at the TCGA cohort. We found a significant difference in LINC00504 expression between Luminal A (n = 231) and basal-like (n = 98) patients (p-value < 0.05). LINC00504 was up-regulated in luminal A samples (p < 0.01) at the TCGA cohort (Figure 4A).

The same comparison between subtypes was made in an independent Brazilian cohort (n = 55) by RT-qPCR. In this case, the samples were classified as triple-negative since the molecular classification information is not available in the patients' clinical reports. In these samples, we also found LINC00504 up-regulation of in luminal A samples (p < 0.01) (Figure 4B).

The next step was to analyze this lncRNA expression's clinical impact on the patient's survival inside each subtype using TCGA cohort follow-up data. The median and quartile values (P25 and P75), were used as cut-offs for the groups, with no difference in survival in luminal A patients (Figure 4C). Notably, basal-like patients with higher median expression of LINC00504 had a higher survival probability (p < 0.05). Therefore, in the basal-like subtype, the high expression of LINC00504 is related to the better prognosis (Figure 4D).

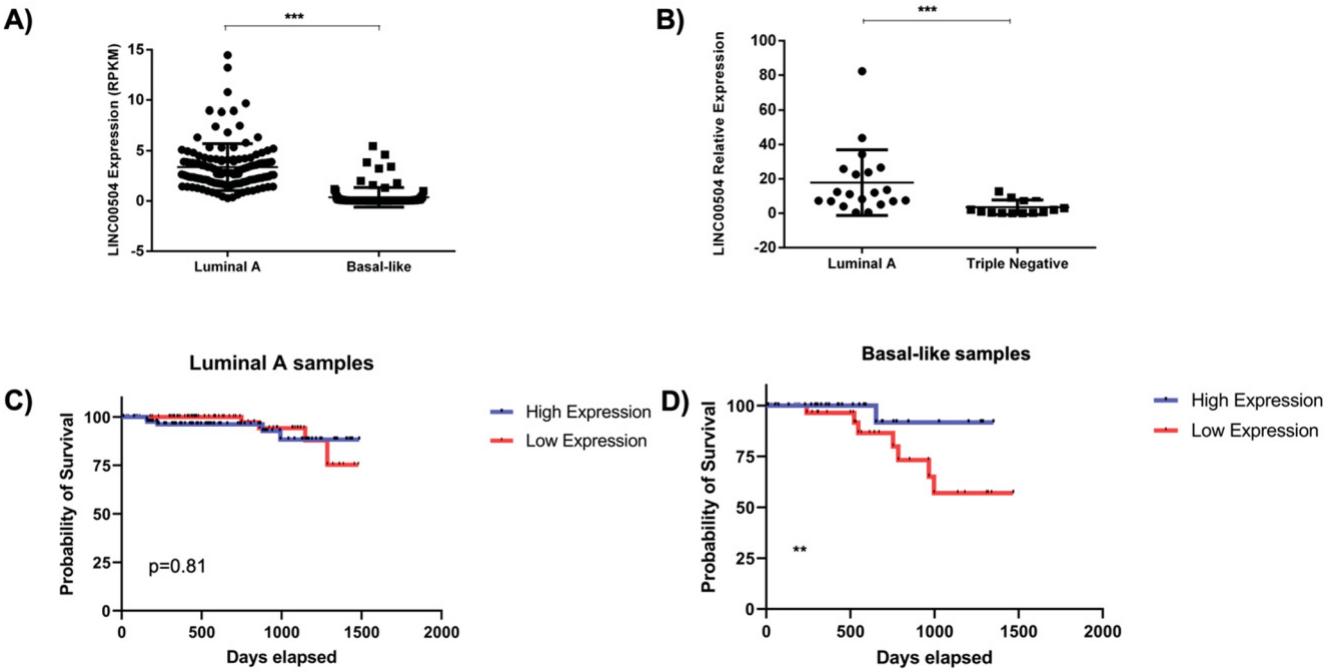


Figure 4. LINC00504 expression and its relationship with patients' survival. (**A**) LINC00504 expression (RPKM) in TCGA BRCA cohort. In total, 231 luminal A and 98 basal-like samples were used, and LINC00504 is up-regulated in luminal A patients; (**B**) LINC00504 relative expression in a Brazilian's patients' cohort. In this cohort, 30 luminal A and 25 triple-negative samples were used, and we observed the same pattern of up-regulation in luminal A samples; (**C**,**D**) Survival analysis considering median LINC00504 expression value as cut-off in Luminal A samples and Basal-like samples, respectively. Red and blue lines refer to patients with low and high expression, respectively. ** $p < 0.05$ and *** $p < 0.01$.

2.5. LINC00504 Is Related to Colony Formation and Cell Viability in Luminal A Cell Lines

To verify the potential biological relevance of LINC00504 in luminal A subtype, we silenced its expression by using the RNA interference technique in two different cell lines, MCF-7 and ZR75-1. These cell lines were chosen because they have a high expression of LINC00504 and are therefore suitable for performing a silencing test (Figure S1). The most efficient result (90% of silencing) was achieved after 48 h after transfection (Figure 5A).

The silencing of LINC00504 decreased the cell viability as measured by Resazurin assay, of the MCF-7 48 h ($p < 0.01$) and 72 h ($p < 0.05$) (Figure 5B) after treatment. For the ZR75-1 cell line, we observed this phenotype 96 h ($p < 0.05$) after treatment (Figure 5C). The number of colonies was significantly reduced for both cell lines after the lncRNA silencing (Figure 5D,E and Figure S2). Apoptosis investigation assays were also carried out on the same strains after the silencing of lncRNA. No statistically significant value was found in this assay.

2.6. LINC00504 Co-Expression Network Has Transcriptional Factors Related to Its Regulation

As previously shown in Figure 3B, the LINC00504 co-expression network contains TFs already related to the luminal A subtype, such as *GATA3*, *FOXA1* and *ESR1*. These TFs were identified in this network, presenting positive mutual information values to LINC00504, thus positively related. With this, it can be assumed that the variation in these RNAs' expressions occur in the same direction, increasing or decreasing together. From the data of mutual information of all the elements identified in this co-expression network, it can also be verified that these genes were the ones that presented higher values of mutual information (Table 1).

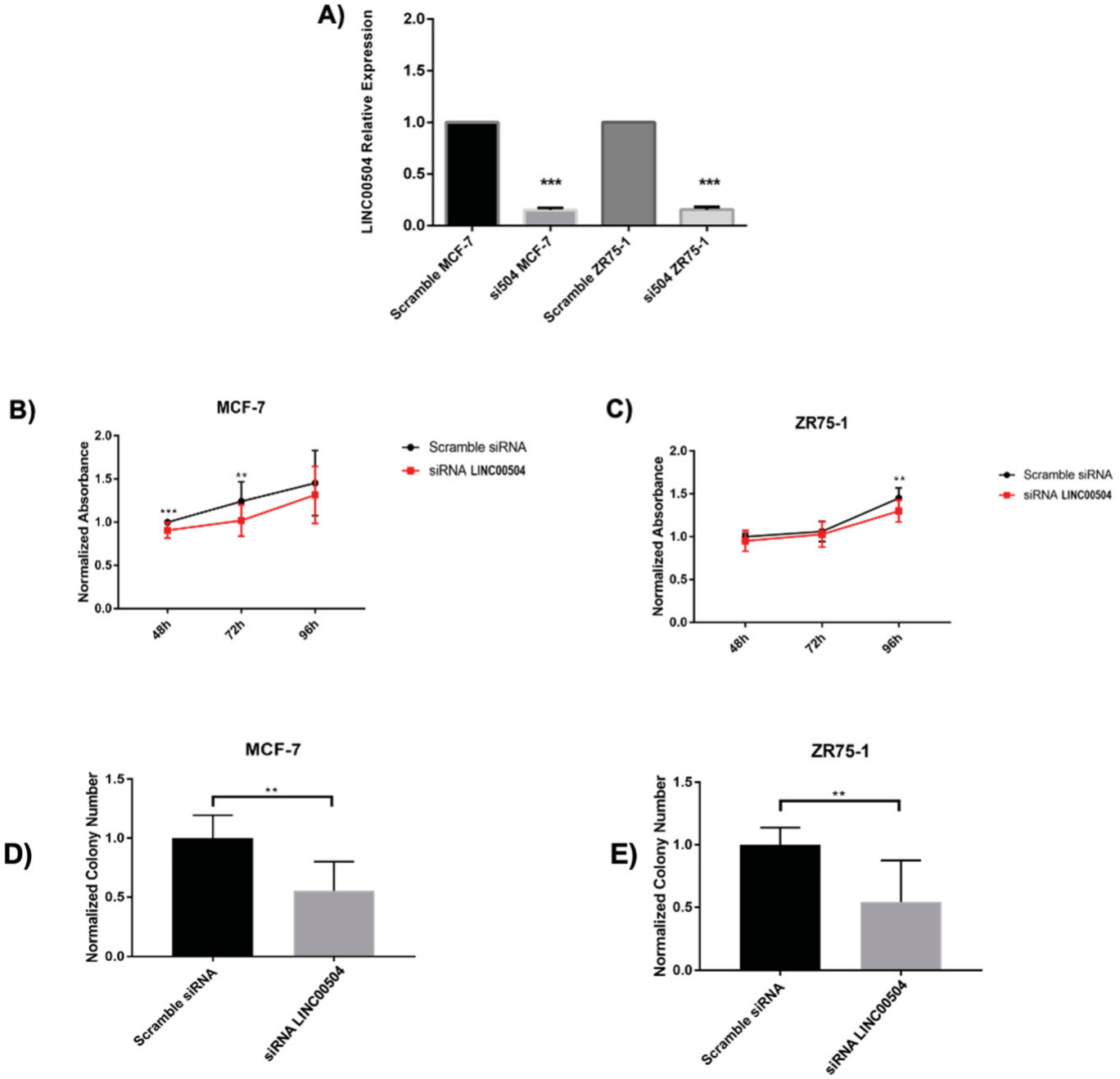


Figure 5. Functional validations of LINC00504 in luminal A cell lines. (**A**) LINC00504 silencing efficiency 48 h after transfection; (**B**) Cell viability in MCF-7 cell line; (**C**) Cell viability in ZR75-1 cell line; (**D**) Colony formation in MCF-7 and (**E**) Colony formation in ZR75-1 cell line; *** $p < 0.01$, ** $p < 0.05$.

Table 1. LINC00504 co-expression network elements detailing.

Network Element	Mutual Information	logFC
CPEB2	0.40	1.85
CPEB2-DT	−0.11	−0.05
ESR1	0.23	6.96
FOXA1	0.20	6.21
GATA3	0.24	4.33

logFC—based on the comparison luminal A x basal-like ($p < 0.05$).

When comparing these RNAs' expression values, considering luminal A x basal-like, it is observed that *ESR1*, *FOXA1,* and *GATA3* are up-regulated in luminal subtype A, justi- fying the positive values of mutual information, since LINC00504 is also up-regulated in luminal A subtype. Based on ChIP-Seq data, made available by the ENCODE project [18,19], it is experimentally evidenced that the GATA3 and FOXA1 bind in the promoter region of LINC00504, possibly regulating its transcription (Figure 6).

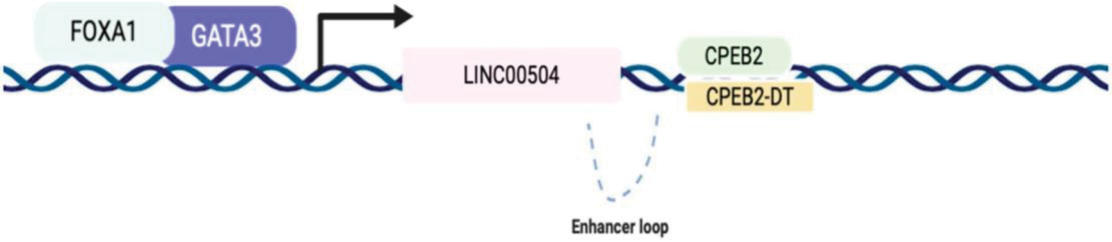


Figure 6. Scheme of transcriptional regulation mediated by GATA3 and FOXA1 in the promoter region of LINC00504. From public data from ChIP-Seq, it was identified that these transcription factors (TF) bind to the promoter region of LINC00504, possibly activating its transcription. These TFs were also identified in the LINC00504 co-expression network.

Another mRNA that stood out for having positive value for mutual information encodes for the Cytoplasmic Polyadenylation Element Binding Protein 2 (*CPEB2*). This protein is required for cell cycle progression, specifically for the transition from metaphase to anaphase [20]. *CPEB2* was shown to promote differentiation and inhibit the epithelial-to- mesenchymal transition in mammary epithelial cells [21]. *CPEB2* elongates the poly(A) tail length of CPE-containing mRNAs, regulating post-transcriptionally mRNAs downstream of steroid hormone signaling, emphasizing the importance of this regulatory molecule in the development of luminal A subtype. Published data from *CPEB2* RNA immunoprecipita- tion [22] identified 169 mRNAs co-immunoprecipitated with *CPEB2* in mammary epithelial cells. We identified 41 of these molecules, also present in LINC00504 co-expression net- work (Supplementary Table S4). Among these mRNAs, *CCND1*, *IGFBP4*, *TIPARP* and *UGCG* belong to the genetic signature "ESTROGEN_RESPONSE_EARLY", identified in the LINC00504 co-expression network enrichment, emphasizing the relevance of this network to the luminal A phenotype.

Through a detailed analysis of the genomic region where LINC00504 is located, we found that the *CPEB2* is located downstream, on the opposite strand. This genomic region is also characterized by the abundance of H3K27Ac epigenetic marks, which are found in regulatory elements, including active enhancers [23]. In fact, when analyzing this region using data deposited on the Genome Browser platform (UCSC), we found a characteristic enhancer loop in the region that comprises these genes, illustrated in Figure 6 It can then be suggested that LINC00504 acts as a trans-enhancer by activating *CPEB2* transcription. This finding requires experimental validation and may represent an important regulatory dependency found in the luminal subtype A.

These results emphasize the robustness of the method used to identify the co-expression networks and identify important lncRNAs. This is the first study to describe the relevance of LINC00504 to breast cancer, and due to the connections established from its network, we show this lncRNA can be important for the luminal A subtype.

3. Discussion

Breast cancer is a heterogeneous disease, encompassing numerous subcategories with differing cellular compositions, molecular alterations, and clinical follow-ups. Due to this heterogeneity, breast cancer classification is an important aspect of therapeutic decision-making [24]. BC can be classified according to histological, immunohistochemical, and molecular profiles. The current BC molecular classifications measure mRNAs expression levels and do not consider non-coding RNAs. However, lncRNAs are gaining prominence

in studying molecular profiles of cancer [25,26] since they present remarkable tissue-specific expression [27].

Many computational methods have been developed to understand and characterize the biology of lncRNAs in physiological and pathological conditions [28,29]. In this work, we propose to use the pipeline contained in the RTN package to search for lncRNAs co-expression networks that provide evidence of their biological relevance in BC. This methodology brings some advantages to others developed with the same objective: Its development was focused on the study of large patient cohorts, and allows the user to set the stringency of the analysis in a stepwise process, including a bootstrap routine designed to remove unstable associations.

We used this strategy to search for lncRNAs co-expression networks with relevance in breast cancer, mainly comparing the luminal A (better prognosis) and basal-like (poor prognosis) subtypes. After using filters, due to the expressive number of results obtained, we call attention to a specific lncRNA (LINC00504) that has not yet been studied in breast cancer and presented exciting results in network analysis.

LINC00504 co-expression network was identified as one with the most significant differences in Differential Enrichment Score (dES) between luminal A and basal-like subtypes. A large positive dES indicates an induced (activated) regulon, while a large negative dES indicates a repressed regulon [30,31].

We identified that LINC00504 is significantly increased in luminal A subtype compared to basal-like. Besides, there was an impact of LINC00504 expression with better prognosis in basal-like patients. Functional studies carried out on luminal A cell lines have shown that LINC00504 down-regulation decreases cell viability and colony formation (Figure 5). Similar to what we observed in breast cancer cell lines, the silencing of LINC00504 reduced cell viability, colony formation and cell migration in colon cancer cell lines [32]. In the same study, LINC00504 silencing also decreased xenograft tumor volume, suggesting that LINC00504 promoted colon cancer development both in vitro and in vivo. In lung cancer cells, LINC00504 expression was higher than in normal cell lines, and its silencing significantly inhibited cell proliferation, colony formation, invasion, and migration and promoted cell apoptosis. Based on these results, LINC00504 was suggested to play an oncogenic role in lung cancer [33]. Another study in lung cancer demonstrated that LINC00504 up-regulation is associated with aggressive progression and poor prognosis in non-small cell lung cancer [34]. The decrease in cell viability and formation of colonies after silencing the LINC00504 was also identified in a study using ovarian cancer cells, indicating that LINC00504 may play an oncogenic role in ovarian cancer cells [35]. These results, in agreement with the results presented in the present study, highlight the relevance of LINC00504 in breast cancer luminal subtype A, suggesting its role as an oncogene in this subtype.

Exploring the data from the co-expression network identified for LINC00504, we highlight here mRNAs that presented the highest positive values of mutual information: *CPEB2*, *ESR1*, *FOXA1* and *GATA3*.

The TFs *FOXA1* and *GATA3*, have been experimentally validated to bind LINC00504 promoter's region, potentially inducing its transcription, according to ENCONDE project public data. The regulation of *ESR1* and LINC00504 has not yet been experimentally validated; however, we know that GATA3 and FOXA1 mediate ESR1 binding to the *cis*-regulatory elements that drive transcription of the *ESR1* target genes. This GATA3 mediation is one of the central components of the ESR1 complex that determines the binding potential and transcriptional targets in breast cancer cells [36]. As suggested by our analyses, these RNAs present positive mutual information, thus varying in the same direction. Therefore, the understanding of the regulatory network of LINC00504 that links these elements together has great biological relevance to the study of luminal A subtype.

According to our analyses, *CPEB2* was the mRNA that presented the highest value of mutual information with LINC00504 (Table 1). CPEB2 regulates the poly(A) tail length of CPE-containing mRNAs, contributing to mammary gland development and luminal breast

carcinogenesis by regulating the translation of mRNAs downstream of steroid hormone signaling [22]. Pascual et al. 2020 [22], using METABRIC and TCGA breast cohorts, verified that *CPEB2* is associated with *ESR1* levels, and high levels of *CPEB2* were associated with worse survival compared to samples with the lowest expression in luminal A patients. On the other hand, ER^- tumors (such as basal-like) do not seem to require *CPEB2*; low levels of *CPEB2* result in reduced survival.

In this same work [22], the authors induced the silencing of *CPEB2* (through the shRNA method) in ZR75-1 cell line, and this depletion significantly decreased cell proliferation in vitro but did not increase apoptosis. Our LINC00504 silencing experiments had similar results to those found in the work of Pascual et al. 2020. This may suggest a mechanism of *CPEB2* and LINC00504 regulation.

CPEB2 is one of the top six genes, together with *ESR1*, with the strongest correlation with ER^+ breast cancer prognosis [36]. Although the cell-of-origin for luminal tumors has not yet been unambiguously identified, these tumors appear to arise from a population of ductal progenitor cells, which have clonogenic capacity and express high levels of markers of mature luminal cells, such as ER, PR, *GATA3*, and *FOXA1* [37,38].

The lncRNA LINC00504 seems to be located in an important regulatory region, which shows high H3K27Ac markers, usually founded in regulatory elements [18]. Through a detailed analysis of this genomic region, it is possible to identify the formation of an enhancer loop housing the LINC00504, CPEB2 and CPEB2-DT genes (Figure 6). This regulatory loop has not yet been experimentally explored and may bring new key regulatory mechanisms in developing luminal subtype.

In summary, the results obtained from the construction of the co-expression network proved to be robust, after considering the 40 networks resulting from the applied filters, by the identification of pathways already related to enriched breast cancer. In addition, we highlight the results found for the LINC00504 co-expression network. In this network, we identified RNAs previously related to each other, and with the luminal subtype A, according to data available in the literature. In addition, functional studies in cells of the luminal subtype A with this lncRNA, have shown to be in agreement with those previously performed in other tumor types. Therefore, we suggest that this pipeline for the construction of regulatory networks can be used in future studies, considering large cohorts of patients, to elucidate relevant biological aspects in the development and progression of the disease.

4. Materials and Methods

4.1. TCGA Data Extraction

The expression data of the lncRNAs were extracted from the TANRIC platform (https://ibl.mdanderson.org/tanric/_design/basic/main.html (accessed on 1 January 2021)). In this database, the TCGA cohort expression data (TCGA, Nature 2012), the BAM files used RPKM to quantify the expression levels of lncRNAs. The mRNA data was extracted from the same TCGA project using the Firebrowse platform (http://firebrowse.org (accessed on 1 January 2021)). The clinical data of the patients in this study were extracted using the cBioPortal platform (https://www.cbioportal.org (accessed on 1 January 2021)).

4.2. RTN Pipeline Implementation

The RTN package is designed for the reconstruction of the transcriptional regulatory network (TRNs) and analysis of regulons using mutual information (MI) [11]. It is implemented by S4 classes in R and extends several methods previously validated for assessing regulons, e.g., MRA [39], GSEA [40], and EVSE [12].

An expression matrix, grouping data from lncRNAs and mRNAs, was organized using the samples that presented the two available data. To select the lncRNAs that would be considered to construct the co-expression networks, we applied a filter under the total number of lncRNAs. Only lncRNAs with detectable expression in at least 90% of the tumor samples used were selected. As a result, 3680 lncRNAs were picked out to reconstruct

the co-expression networks using the RTN package. We used a total of 1000 permutations to predict these networks, and this analysis was performed considering three levels of significance, $p < 10^{-6}$, $p < 10^{-7}$, and $p < 10^{-8}$ for better visualization of the most significant co-expression networks. To ensure a higher level of stringency of the analysis, we used the co-expression networks computed at the significance level of $p < 10^{-8}$.

The regulon's activity heatmap was also performed using RTN package. Ranked differential Enrichment Score (dES) plot for the BRCA cohort and status of key attributes plot were constructed using RTN Survival package [41].

4.3. Regulon Filtering in Breast Cancer Molecular Subtypes

After the reconstruction of 3,680 co-expression networks, we exclude those that ex- hibit size less than 15 targets and those who had no Differential Enrichment Score (dES) − 0.5 < dES > 0.5, thus, avoiding the selection of networks with small informative value. This selection resulted in a final number of 84 co-expression networks.

These 84 co-expression networks were then stratified according to breast cancer molecular subtype: Basal-like, HER2-enriched, Luminal A, and Luminal B. Normal-like samples were withdrawn from the analysis because of the small sample number ($n = 7$). The non-parametric Kruskal-Wallis test and post hoc Dunn's test were used to verify the difference in the variance of dES comparing the breast cancer subtypes for the 84 regulatory networks selected. For additional analysis, only networks with p-value < 0.05 in both tests were considered.

At this point, we focus on the subsequent analyzes on the co-expression networks with the greatest difference of dES value among the subtypes of greatest differences in prognosis, luminal A and Basal-like. With this criterion, we aim to select co-expression networks that are differentially regulated between these subtypes. This selection resulted in 40 co-expression networks evaluated by available literature data, regarding the lncRNAs used to construct the co-expression networks and by the identification of genes previously described as relevant in mammary carcinogenesis.

4.4. Cancer-Related Processes Enrichment Analysis

The enrichment analysis of the co-expression networks was performed using Molec- ular Signatures Database (MSigDB). Within this database, we use the genetic signatures contained in the collection called "Hallmarks". This group are coherently expressed signa- tures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes. In total, 50 genetic signatures of the "Hallmarks" collection are de- posited and these are available at https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H (accessed on 1 January 2021). The graphical representation of the enriched hallmarks was made using the ggplot2 package.

4.5. Breast Cancer Tissue Sample Characterization

A total of 55 breast tissue samples from different patients, predominantly of European descent, who had been diagnosed with breast cancer were collected during primary surgery at Hospital Nossa Senhora Das Graças (HNSG) from Curitiba. Clinical and histopathologic data of the patients were collected directly from the medical records in a coded manner without patient identifiers. The samples were classified according to their immunohis- tochemical profile according to Goldhirsch et al. [42]. In total, 30 samples classified as Luminal A and 25 as Triple negative were used. This study was approved by the Brazilian Commission of Ethics in Research (CONEP) under the number CAAE 67400917.3.0000.55 following Brazilian Federal laws in 20/02/2003. All participants signed informed written consent following the principles of the Declaration of Helsinki.

4.6. RNA Isolation and Expression Quantification

RNA isolation of the fresh tumor samples and cell lines was performed using Re- coverAll Total Nucleic Acid Isolation Kit (Invitrogen, Carlsbad, CA, USA) and TRIzol®

Reagent (Invitrogen, Carlsbad, CA, USA) respectively, according to the manufacturer's protocol. After the isolation, 1000 ng of RNA (each sample) were treated with DNAse I (Invitrogen, Carlsbad, CA, USA) and converted into cDNA using the SuperScript III enzyme (Invitrogen, Carlsbad, CA, USA) using random hemaxers, and following the man- ufacturer's protocol. Real-time PCR was performed using the Viia-7 Sequence Detection Systems (Applied Biosystems, San Francisco, CA, USA) with Power SYBR Green PCR Master Mix (Applied Biosystems). For each reaction, we used RT- as a control and as a normalizer between the PCR reactions a cDNA pool of breast cancer tumor lines. The gene expression was determined using the $2^{-\Delta\Delta Ct}$ method [43], using β-glucuronidase (*GUS*β) and actin-beta (*ACTB*) expression as normalizers. The reactions were performed in triplicate, and negative control was included for each set of primers in each batch. The primer sequences are organized in Supplementary Table S1.

4.7. Cell Culture and Growth Conditions

MCF-7 and ZR-75-1cells were cultured in RPMI medium (Gibco, Grand Island, NY, USA) supplemented with 10% FBS (Gibco) and 1% Penicillin-Streptomycin and maintained a humidified incubator with 5% CO_2.

4.8. siRNA Treatment

A total of 6×10^5 MCF-7 and ZR75-1 cells were seeded onto six-well plates, reaching a total of 50–70% confluence. The cells were transfected with 125 nM siRNA scramble and siRNA LINC00504 (from the catalog of the company ThermoFisher, Waltham, MA, USA), number 4392420). Cells were reverse transfected using Lipofectamine 2000 reagent (Invitrogen), according to the manufacturer's protocol and evaluated after 24, 48, and 72 h.

4.9. Cell Viability

For cell viability assay, MCF-7 and ZR75-1 cells treated with siRNA scramble and siLINC00504 were plated in 96 well plates. A total of 500 cells were seeded in each well. The cell viability was evaluated after 48, 72, and 96 h of transfection. On each day, the media were replaced with 100 µL of fresh media + 10 µL Resazurin dye (7-hydroxy-3H- phenoxazin-3-one 10-oxide) at concentration of 0.1 mg/mL and incubated for 4 h, and read at 600 ηm and 570 ηm using TECAN Infinite® 200 PRO system. The experiment was made in triplicate.

4.10. Colony-Forming Assay

The cells were seeded in a 6-well plate (1000 cells each well) and grew for 25 days at 37 °C in a 5% CO_2 humidified incubator after treatment with siLINC00504 or siRNA scramble. The cells were fixed with 100% methanol for 20 min, stained with 1% crystal violet for 5 min, and washed with water until excess dye is removed (every step was performed at room temperature). The experiment was carried out in triplicate.

4.11. Apoptosis

The apoptosis ratio was analyzed using the Annexin V-FITC Apoptosis Detection Kit. Forty-eight hours after transfection, MCF-7 and ZR75-1 cells were harvested and resuspended in a binding buffer containing Annexin V-FITC and P.I. (propidium iodide) according to its instructions. The samples were analyzed by flow cytometry (B.D. Biosciences, San Jose, CA, USA). Cells were discriminated into viable cells, necrotic cells, early apoptosis, and later apoptosis cells using the BD FACSVantage™ cytofluorimeter (B.D. Biosciences, USA), and then the percentages of apoptotic cells from each group were compared. Data analysis was performed using FlowJo software (v.10). The experiment was performed in triplicate.

4.12. Statistical Analysis

Data are presented as mean S.D. Statistical analysis of the data was performed by Student's *t*-test using R (version 3.6.3) (http:///www.r-project.org/ (accessed on 1 January 2021)). *p*-values of ≤0.05 were considered statistically significant.

Supplementary Materials: The following are available online at https://www.mdpi.com/1422-0067/22/5/2420/s1. Table S1: Primers sequences utilized in RT-qPCR; Table S2: 84 LncRNAs co- expression networks after dES filtering; Table S3: 40 LncRNAs co-expression networks after luminal A and basal-like extratification filtering; Table S4: Common mRNAs identified in CPEB2 coimunopre- ciptation and LINC00504 co-expression network; Figure S1: LINC00504 relative expression in the tested breast cancer cell lines; Figure S2: Representative figure of colony formation assay in MCF-7 and ZR75-1 cell lines.

Author Contributions: Conceptualization, C.M., M.A.A.C., D.F.G., and J.C.d.O.; methodology, C.M., C.S.G., S.T., M.A.A.C.; samples acquisition and classification, R.S.L., C.A.U., E.M.S.F.R.; validation, C.M., E.P.Z., K.B.P.; formal analysis, C.M., C.S.G., S.T., M.A.A.C., E.P.Z., K.B.P., D.F.G., and J.C.d.O.; writing–original draft preparation, C.M., D.F.G., J.C.d.O.; supervision: M.A.A.C., D.F.G., and J.C.d.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Public Research Agencies CAPES (001- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) processo 153771/2018-6.

Institutional Review Board Statement: This study was approved by the Brazilian Commission of Ethics in Research (CONEP) under the number CAAE 67400917.3.0000.55 following Brazilian Federal laws in 20/02/2003. All participants signed informed written consent following the principles of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is contained within the article or Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hombach, S.; Kretz, M. Non-coding RNAs: Classification, Biology and Functioning. *Adv. Exp. Med. Biol.* **2016**, *937*, 3–17. [CrossRef]
2. Clark, M.B.; Mercer, T.R.; Bussotti, G.; Leonardi, T.; Haynes, K.R.; Crawford, J.; Brunck, M.E.; Le Cao, K.A.; Thomas, G.P.; Chen, W.Y.; et al. Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods* **2015**, *12*, 339. [CrossRef]
3. Chan, J.J.; Tay, Y. Noncoding RNA:RNA Regulatory Networks in Cancer. *Int. J. Mol. Sci.* **2018**, *19*, 1310. [CrossRef]
4. Bartonicek, N.; Maag, J.L.; Dinger, M.E. Long noncoding RNAs in cancer: Mechanisms of action and technological advancements. *Mol. Cancer* **2016**, *15*, 43. [CrossRef] [PubMed]
5. Chi, Y.; Wang, D.; Wang, J.; Yu, W.; Yang, J. Long Non-Coding RNA in the Pathogenesis of Cancers. *Cells* **2019**, *8*, 1015. [CrossRef]
6. Ghafouri-Fard, S.; Taheri, M. Maternally expressed gene 3 (MEG3): A tumor suppressor long non coding RNA. *Biomed. Pharmacother.* **2019**, *118*, 109129. [CrossRef] [PubMed]
7. Qian, Y.; Shi, L.; Luo, Z. Long Non-coding RNAs in Cancer: Implications for Diagnosis, Prognosis, and Therapy. *Front Med.* **2020**, *7*, 612393. [CrossRef]
8. Jalali, S.; Kapoor, S.; Sivadas, A.; Bhartiya, D.; Scaria, V. Computational approaches towards understanding human long noncoding RNA biology. *Bioinformatics* **2015**, *31*, 2241–2251. [CrossRef]
9. Liao, Q.; Liu, C.; Yuan, X.; Kang, S.; Miao, R.; Xiao, H.; Zhao, G.; Luo, H.; Bu, D.; Zhao, H.; et al. Large-scale prediction of long noncoding RNA functions in a coding-non-coding gene coexpression network. *Nucleic Acids Res.* **2011**, *39*, 3864–3878. [CrossRef] [PubMed]
10. Necsulea, A.; Soumillon, M.; Warnefors, M.; Liechti, A.; Daish, T.; Zeller, U.; Baker, J.C.; Grützner, F.; Kaessmann, H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **2014**, *505*, 635–640. [CrossRef]
11. Fletcher, M.N.; Castro, M.A.; Wang, X.; De Santiago, I.; O'Reilly, M.; Chin, S.F.; Rueda, O.M.; Caldas, C.; Ponder, B.A.; Markowetz, F.; et al. Master Regulators of FGFR2 Signalling and Breast Cancer Risk. *Nat. Commun.* **2013**, *4*, 2464. [CrossRef]
12. Castro, M.A.; de Santiago, I.; Campbell, T.M.; Vaughn, C.; Hickey, T.E.; Ross, E.; Tilley, W.D.; Markowetz, F.; Ponder, B.A.; Meyer, K.B. Regulators of Genetic Risk of Breast Cancer Identified by Integrative Network Analysis. *Nat. Genet.* **2016**, *48*, 12–21. [CrossRef] [PubMed]
13. Perou, C.M.; Sørlie, T.; Eisen, M.B.; Van De Rijn, M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslen, L.A.; et al. Molecular portraits of human breast tumours. *Nature* **2000**, *406*, 747–752. [CrossRef]
14. Yersal, O.; Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J. Clin. Oncol.* **2014**, *5*, 412–424. [CrossRef]
15. Beniey, M.; Haque, T.; Hassan, S. Translating the role of PARP inhibitors in triple-negative breast cancer. *Oncoscience* **2019**, *6*, 287–288. [CrossRef]
16. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries published correction. *CA Cancer J. Clin.* **2018**, *68*, 394–424, Erratum in **2020**, *70*, 313. [CrossRef] [PubMed]
17. Polyak, K. Breast cancer: Origins and evolution. *J. Clin. Investig.* **2007**, *117*, 3155–3163. [CrossRef]
18. Davis, C.A.; Hitz, B.C.; Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Gabdank, I.; Hilton, J.A.; Jain, K.; Baymuradov, U.K.; Narayanan, A.K.; et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **2018**, *46*, D794–D801. [CrossRef]
19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef]
20. Giangarrà, V.; Igea, A.; Castellazzi, C.L.; Bava, F.A.; Mendez, R. Global Analysis of CPEBs Reveals Sequential and Non-Redundant Functions in Mitotic Cell Cycle. *PLoS ONE* **2015**, *10*, e0138794. [CrossRef] [PubMed]
21. Grudzien-Nogalska, E.; Reed, B.C.; Rhoads, R.E. CPEB1 promotes differentiation and suppresses EMT in mammary epithelial cells. *J. Cell Sci.* **2014**, *127 Pt 10*, 2326–2338. [CrossRef]
22. Pascual, R.; Martín, J.; Salvador, F.; Reina, O.; Chanes, V.; Millanes-Romero, A.; Suñer, C.; Fernández-Miranda, G.; Bartomeu, A.; Huang, Y.S.; et al. The RNA binding protein CPEB2 regulates hormone sensing in mammary gland development and luminal breast cancer. *Sci. Adv.* **2020**, *6*, eaax3868. [CrossRef]
23. Spicuglia, S.; Vanhille, L. Chromatin signatures of active enhancers. *Nucleus* **2012**, *3*, 126–131. [CrossRef] [PubMed]
24. Al-Thoubaity, F.K. Molecular classification of breast cancer: A retrospective cohort study. *Ann. Med. Surg.* **2019**, *49*, 44–48. [CrossRef] [PubMed]
25. Ung, M.; Schaafsma, E.; Mattox, D.; Wang, G.L.; Cheng, C. Pan-cancer systematic identification of lncRNAs associated with cancer prognosis. *PeerJ* **2020**, *8*, e8797. [CrossRef] [PubMed]
26. Huang, R.; Zhou, L.; Chi, Y.; Wu, H.; Shi, L. LncRNA profile study reveals a seven-lncRNA signature predicts the prognosis of patients with colorectal cancer. *Biomark Res.* **2020**, *8*, 8. [CrossRef] [PubMed]
27. Gu, W.; Gao, T.; Sun, Y.; Zheng, X.; Wang, J.; Ma, J.; Hu, X.; Li, J.; Hu, M. LncRNA expression profile reveals the potential role of lncRNAs in gastric carcinogenesis. *Cancer Biomark.* **2015**, *15*, 249–258. [CrossRef]
28. Jiang, C.; Li, Y.; Zhao, Z.; Lu, J.; Chen, H.; Ding, N.; Wang, G.; Xu, J.; Li, X. Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* **2016**, *7*, 7120–7133. [CrossRef]
29. Yan, C.; Zhang, Z.; Bao, S.; Hou, P.; Zhou, M.; Xu, C.; Sun, J. Computational Methods and Applications for Identifying Disease-Associated lncRNAs as Potential Biomarkers and Therapeutic Targets. *Mol. Ther. Nucleic Acids* **2020**, *21*, 156–171. [CrossRef]
30. Campbell, T.M.; Castro, M.A.; de Oliveira, K.G.; Ponder, B.A.; Meyer, K.B. ERα Binding by Transcription Factors Nfib and Ybx1 Enables Fgfr2 Signaling to Modulate Estrogen Responsiveness in Breast Cancer. *Cancer Res.* **2018**, *78*, 410–421. [CrossRef]
31. Campbell, T.M.; Castro, M.A.; de Santiago, I.; Fletcher, M.N.; Halim, S.; Prathalingam, R.; Ponder, B.A.; Meyer, K.B. FGFR2 Risk Snps Confer Breast Cancer Risk by Augmenting Oestrogen Responsiveness. *Carcinogenesis* **2016**, *37*, 741–750. [CrossRef]
32. Feng, J.; Ma, J.; Liu, S.; Wang, J.; Chen, Y. A noncoding RNA LINC00504 interacts with c-Myc to regulate tumor metabolism in colon cancer. *J. Cell Biochem.* **2019**, *120*, 14725–14734. [CrossRef]
33. Zhang, Z. Silencing LINC00504 inhibits cell proliferation, invasion as well as migration and promotes cell apoptosis in lung cancer cells via upregulating miR-876-3p. *Cytotechnology* **2020**. [CrossRef]

34. Ma, H.P.; Wang, L.X.; Li, W.; Guo, H.H.; Wu, Y.; Li, X.Y. Upregulation of LINC00504 is associated with aggressive progression and poor prognosis in non-small cell lung cancer. *Eur. Rev. Med. Pharmacol. Sci.* **2020**, *24*, 699–703. [CrossRef]
35. Liu, Y.; He, X.; Chen, Y.; Cao, D. Long noncoding RNA LINC00504 regulates the Warburg effect in ovarian cancer through inhibition of miR-1244. *Mol. Cell Biochem.* **2020**, *464*, 39–50. [CrossRef]
36. Theodorou, V.; Stark, R.; Menon, S.; Carroll, J.S. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* **2013**, *23*, 12–22. [CrossRef] [PubMed]
37. Xiao, B.; Hang, J.; Lei, T.; He, Y.; Kuang, Z.; Wang, L.; Chen, L.; He, J.; Zhang, W.; Liao, Y.; et al. Identification of key genes relevant to the prognosis of ER-positive and ER-negative breast cancer based on a prognostic prediction system. *Mol. Biol. Rep.* **2019**, *46*, 2111–2119. [CrossRef] [PubMed]
38. Shehata, M.; Teschendorff, A.; Sharp, G.; Novcic, N.; Russell, I.A.; Avril, S.; Prater, M.; Eirew, P.; Caldas, C.; Watson, C.J.; et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **2012**, *14*, R134. [CrossRef] [PubMed]
39. Carro, M.S.; Lim, W.K.; Alvarez, M.J.; Bollo, R.J.; Zhao, X.; Snyder, E.Y.; Sulman, E.P.; Anne, S.L.; Doetsch, F.; Colman, H.; et al. The Transcriptional Network for Mesenchymal Transformation of Brain Tumours. *Nature* **2010**, *463*, 318–325. [CrossRef]
40. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2015**, *102*, 15545–15550. [CrossRef]
41. Groeneveld, C.; Chagas, V.; Jones, S.; Robertson, G.; Ponder, B.; Meyer, K.; Castro, M. RTNsurvival: An R/Bioconductor package for regulatory network survival analysis. *Bioinformatics* **2019**, btz229. [CrossRef] [PubMed]
42. Goldhirsch, A.; Winer, E.P.; Coates, A.S.; Gelber, R.D.; Piccart-Gebhart, M.; Thürlimann, B.; Senn, H.J.; Panel Members. Personaliz- ing the treatment of women with early breast cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann. Oncol.* **2013**, *24*, 2206–2223. [CrossRef] [PubMed]
43. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]

8. CAPÍTULO III

frontiers in Oncology

OPEN ACCESS

Edited by:
Wenwen Zhang, Nanjing Medical University, China

Reviewed by: Liuzhi Zhou, Zhejiang University, China Ryuichi Sugino, Saitama Cancer Center, Japan

*Correspondence: Jaqueline Carvalho de Oliveira jaqueline.carvalho@ufpr.br

†These authors have contributed equally to this work

‡These authors have jointly directed this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: 07 April 2021
Accepted: 06 May 2021
Published: 31 May 2021

Citation:
Mathias C, Muzzi JCD, Antunes BB, Gradia DF, Castro MAA and Carvalho de Oliveira J (2021) Unraveling Immune-Related lncRNAs in Breast Cancer Molecular Subtypes. Front. Oncol. 11:692170. doi: 10.3389/fonc.2021.692170

ORIGINAL RESEARCH

Unraveling Immune-Related lncRNAs in Breast Cancer Molecular Subtypes

Carolina Mathias [1†], João Carlos Degraf Muzzi [2,3,4†], Bruna Borba Antunes [1,2], Daniela F. Gradia [1], Mauro A. A. Castro [2‡] and Jaqueline Carvalho de Oliveira [1*‡]

[1] Department of Genetics, Federal University of Parana, Post-graduation Program in Genetics, Curitiba, Brazil,
[2] Bioinformatics and Systems Biology Lab, Federal University of Parana (UFPR), Polytechnic Center, Curitiba, Brazil,
[3] Immunochemistry Laboratory (LIMQ), Federal University of Parana, Post-graduation Program in Microbiology, Parasitology and Pathology, Curitiba, Brazil, [4] Instituto de Pesquisa Pelé Pequeno Príncipe, Oncology Division, Curitiba, Brazil

Breast cancer (BRCA) is the most leading cause of cancer worldwide. It is a heterogeneous disease with at least five molecular subtypes including luminal A, luminal B, basal-like, HER2-enriched, and normal-like. These five molecular subtypes are usually stratified according to their mRNA profile patterns; however, ncRNAs are increasingly being used for this purpose. Among the ncRNAs class, the long non-coding RNAs (lncRNAs) are molecules with more than 200 nucleotides with versatile regulatory roles; and high tissue-specific expression profiles. The heterogeneity of BRCA can also be reflected regarding tumor microenvironment immune cells composition, which can directly impact a patient's prognosis and therapy response. Using BRCA immunogenomics data from a previous study, we propose here a bioinformatics approach to include lncRNAs complexity in BRCA molecular and immune subtype. RNA-seq data from The Cancer Genome Atlas (TCGA) BRCA cohort was analyzed, and signal-to-noise ratio metrics were applied to create these subtype-specific signatures. Five immune-related signatures were generated with approximately ten specific lncRNAs, which were then functionally analyzed
using GSEA enrichment and survival analysis. We highlighted here some lncRNAs in each subtype. LINC01871 is related to immune response activation and favorable overall survival in basal-like samples;

EBLN3P is related to immune response suppression and progression in luminal B, MEG3, XXYLT1-AS2, and LINC02613 were related with immune response activation in luminal A, HER2-enriched and normal-like subtypes, respectively. In this way, we emphasize the need to know better the role of lncRNAs as regulators of immune response to provide new perspectives regarding diagnosis, prognosis and therapeutical targets in BRCA molecular subtypes.

Keywords: immune response, MEG3, LINC01871, EBLN3P, LINC02613, XXYLT1-AS2

INTRODUCTION

Breast cancer (BRCA) is a molecular and histological heterogeneous disease with at least five intrinsic molecular subtypes (1, 2). Based on gene expression, BRCA can be mainly classified into luminal A (LumA), luminal B (LumB), HER2-enriched (Her2), basal-like (Basal), and normal-like (Normal) (3, 4). These subtypes have a distinct prognosis and also differ according to therapeutic response. LumA and LumB tumors respond well to hormonal interventions, while HER2+ tumors respond effectively when anti-HER2 therapy is used (5). Basal tumors are very aggressive and associated with the shortest survival times, with no current molecular-based targeted therapies available (6).

Immunotherapy brought a new line of action in cancer care; however its response varies across cancer types and patients. The immune system response in the tumor microenvironment may help to guide immunotherapy drug discovery and clinical decisions (7). In general, tumors more responsive to immune checkpoint inhibitors are related to high levels of leukocyte fraction in the tumor microenvironment (8). Besides gene expression differences in BRCA molecular subtypes, they differ significantly concerning the composition of cells that form the tumor microenvironment, especially the immune system's cells. A substantial proportion of natural killer cells and neutrophils have been found in luminal tumors. In contrast, in these tumors, cytotoxic T cells (T CD8+) and naïve and memory T cells are found less frequently. In BRCA Basal tumors, T regs, associated macrophages 2, and activated mast cells form a significant portion of the immune infiltrate cells. The immune infiltrate composition is not widely described in the Her2 subtype. It is known that it is mainly formed by dendritic cells, mast cells, gd T lymphocytes, T regs and neutrophils (9).

A landscape of tumors' immune microenvironment was characterized from immunogenomics data by Thorsson and colleagues (8). In this study, using an integrated analysis, they could classify solid tumors (from The Cancer Genome Atlas) into six major immune subtypes, which they called C1-C6. These subtypes have distinct immune signature sets, which could also be related to prognosis. C1 (wound healing) exhibited elevated expression of angiogenic genes, a high proliferation rate, and a Th2 cell bias to the adaptive immune infiltrate. C2 (IFN-g dominant) had a strong T CD8+ signal, the greatest TCR diversity, and a high proliferation rate. C3 (inflammatory) was the subtype that presented high Th17 and Th1 genes and low to moderate tumor cell proliferation. C4 (lymphocyte depleted) displayed a more prominent macrophage signature with Th1 suppressed and high M2 response. C5 (immunologically quiet) was enriched by brain tumors and exhibited the lowest lymphocyte and most increased macrophage responses. Finally, C6 (TGF-b dominant) displayed the highest TGF-b signature and a high lymphocytic infiltrate with an even distribution of type I and type II T cells (8).

According to this approach, BRCA could be classified into five subtypes (C1, C2, C3, C4 and C6), being C2 (n=345) the most representative subtype, followed by C1 (n=320). Immune subtypes were also described according to BRCA molecular subtypes, and as expected, the subtypes varied significantly according to these immune groups. For example, LumA was more representative of the C1 subtype, while Basal samples of C2 (8).

Gene expression sets based on mRNAs were used for the classification and determination of molecular and immune subtypes. However, it is already known that the most abundant part of the human genome is not translated into proteins. These transcripts are organized in a class called "non-coding RNAs." Non-coding RNAs can be classified into two major categories based on their size: small non-coding RNAs (<200 nucleotides) and long non-coding RNAs (>200 nucleotides) (10). LncRNAs are usually transcribed by RNA polymerase II, polyadenylated, and capped. They exhibit high tissue specificity and great regulatory versatility, acting at different gene expression regulation levels (11, 12).

Due to its high tissue specificity, lncRNAs can be evaluated as potential disease biomarkers, including BRCA (13–15). Based on this, we looked for molecular subtype specific lncRNAs signatures that could help differentiate the immune profiles described in Thorsson (8). These lncRNAs were also analyzed if impact the patient's overall survival and progression free interval and were also investigated in differential expression and enrichment analysis to explore other possible biological roles of these lncRNAs in BRCA molecular and immune subtypes.

MATERIAL AND METHODS

Breast Cancer Immune Data

BRCA molecular and immune subtypes, leukocyte fraction, and survival information were downloaded from Thorsson et al. (8) Supplementary Material. According to samples' barcode expression and immune type, data were integrated to perform the analysis. In Supplementary Table 1, we organized data according to breast cancer molecular and immune subtypes.

Breast Cancer RNA-Seq Data

Log2 normalized FPKM RNA-Seq data from The Cancer Genome Atlas (TCGA) breast cancer cohort was downloaded from XenaBrowser (https://xenabrowser.net/datapages/), and primary tumor samples were selected and merged with Thorsson et al. (8) master table using patients' barcode. The lncRNAs and protein-coding gene expression profiles were filtered from the RNA-Seq data using the R package biomaRt v 2.46.3 (16). For lncRNAs, when available, HGNC Symbol was used; otherwise, Ensembl gene name was used. All Ensembl and HGNC Symbols from lncRNAs used in this study are available in Supplementary Table 2. The non-tumoral samples were selected based on TCGA barcodes ending with 11A or 11B. The molecular BRCA subtypes were defined as described in Thorsson et al. (8) Supplementary Material, based on PAM50. The expression profiles of immunomodulatory genes listed (8) (https://www.cell.com/cms/10.1016/j.immuni.2018.03.023/attachment/8d3ffc74-4db4-4531-a4ad-389dfc8bb7ec/mmc7.xlsx) previously were obtained from the gene expression matrix. Of the 75 immune modulators, only one (C10orf54) was not

found in the expression matrix. For heatmap construction using *ComplexHeatmap* package (1), samples were displayed in columns and genes in the rows. Column-wise z-score was calculated for gene expression values, and maximum and minimum values were limited to +2 and −2 standard deviations, respectively. Samples were clustered within each BRCA subtype.

The BRCA lncRNAs expressions were further filtered above the first quartile for lncRNA expression sum, which means lncRNAs with expression sum above 8.04 log2 FPKM in the whole BRCA cohort (1,060 samples). Signal-to-noise ratio (SNR) was calculated for each molecular subgroup individually as follow:

$$SNR = \frac{\mu 1 - \mu 2}{\sigma 1 + \sigma 2}$$

being m_1 the mean of lncRNA expression in the group analyzed and m_2 the mean of lncRNA expression in the patients out of the group analyzed. s refers to the standard deviation from the respective groups mentioned. We selected the lncRNAs above the SNR 0.95 quantile for each BRCA molecular subtype, that means the lncRNAs with higher expression in the subtype analyzed compared with the rest of the cohort. Venn diagram was constructed using *InteractiVenn* (http://www.interactivenn.net/) demonstrating the intersection of lncRNA between groups. Then, we calculated the SNR within each molecular subgroup based on the immune subtypes described previously (8). We considered for the analysis the immune groups with more than five patients in each molecular subtype. In this way, Basal and Her2 samples were divided into C1 and C2 subtypes; LumA and Normal into C1, C2, C3, C4, and C6 subtypes and LumB into C1, C2, C3, and C4 subtypes. After absolute SNR sum calculation, we selected lncRNAs considered in 0.98 quantile, which means the lncRNAs with the most significant variation within the immune subtypes for each molecular subgroup. For histogram construction, the absolute SNR sum was scaled using z-score. The 53 lncRNAs selected from this analysis are presented in Supplementary Table 3.

Leucocyte Fraction Correlation

The lncRNAs expression was used to calculate Spearman correlation with the leucocyte fraction observed in each BRCA subtype, and the p-value was calculated with AS 89 algorithm (18) using cor. test function from stats R package v.4.0.4 (19) and adjusted by False Discovery Rate (FDR) method.

Survival Analysis

Survival analysis was firstly performed using the *coxph* function available in the survival R package v3.2-10 (20) based on lncRNA expression for both Overall Survival (OS) and Progression-Free Interval (PFI) for each lncRNA individually in its respective BRCA molecular subgroup. The univariate Cox p-value for each lncRNA was calculated and adjusted by the FDR method; values below 10% FDR level were considered significant.

One lncRNA was selected on each BRCA subtype based on Cox results or Leukocyte Fraction correlation for further analysis. Kaplan-Meier was calculated, and patients were divided by the median lncRNA expression value in High Expression and Low Expression; p-value was calculated by log- rank test.

Enrichment Analysis

The protein-coding genes expression profile was filtered for each molecular subgroup by patients' barcode, and genes with zero sum expression were removed. Patients were divided by the median lncRNA expression value, and High versus Low Expression groups were used for SNR calculus. Genes were ordered by SNR value, and gene set enrichment analysis (GSEA) was inferred using fgsea R package v1.16.0 (21) with MSigDb v7.2 Hallmarks gene sets (22, 23) for 10,000 permutations.

RESULTS

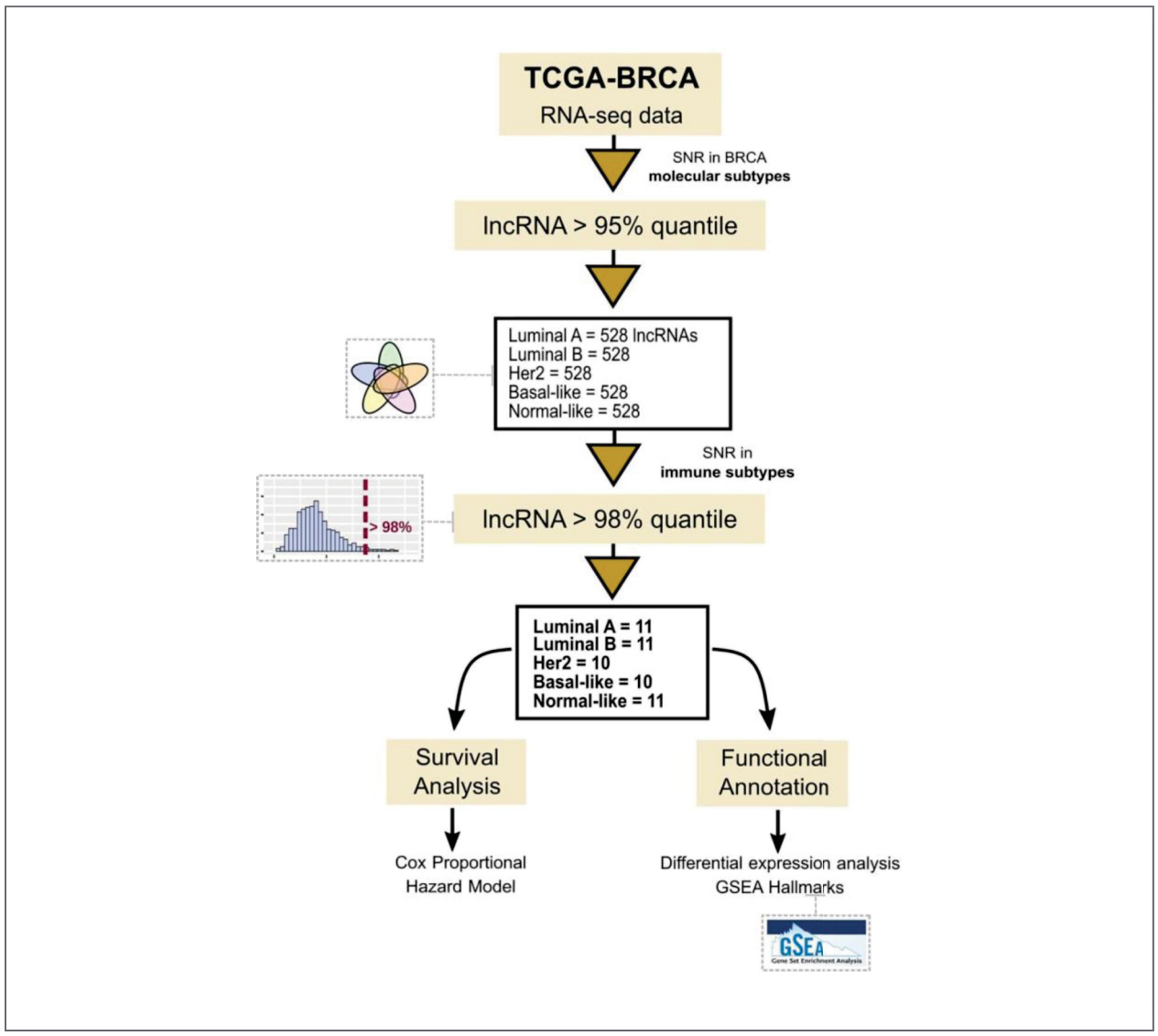
Immune Modulator Genes Expression Demonstrates Distinct Patterns Within BRCA Molecular Subtypes

Based on the gene expression profile related to immune response, Throsson et al. (8) analyzed over 10,000 TCGA samples, from which 1,087 were from BRCA samples and clustered them in six immune subtypes. We merged the barcodes with the gene expression matrix downloaded from XenaBrowser, remaining 1,060 BRCA primary tumor samples. The molecular classification as presented in Thorsson et al. (8) master table as TCGA Subtype was used, dividing the samples into five groups, LumA (n=499), LumB (n = 184), Basal (n=169), Normal (n=136) and Her2 (n=72). In Supplementary Table 1, we represented the number of samples according to each molecular subgroup's immune subtype.

Supplementary Figure 1 shows a distinct pattern of immune modulators gene expression in all molecular subtypes. In general, most genes seem to be upregulated in the same samples (columns of the heatmap) independent of its classification as an inhibitor or stimulator of the immune system. Overall, the immune activation seen in the gene expression follows the rising in Leukocyte Fraction and tends to group the C2 immune subtype. Simultaneously, the inverse is observed for a low expression pattern associated with low Leukocyte Fraction and C1 immune subtype. Basal and Her2 demonstrate a more apparent separation of C1 and C2. These two immune subtypes are the major representatives in these molecular groups; for instance, C1 and C2 represent 95.9% and 93.1% of all Basal and Her2 samples, respectively (Supplementary Table 1).

SNR Highlights Specific lncRNAs for Each BRCA Molecular Subgroups Related to Immune Subtypes Distinction

The first step of our strategy is to search for lncRNAs that could be associated with immune subtypes in BRCA patients. For this, lncRNAs with greater distinct expression patterns among the five molecular subgroups were selected (Figure 1). After selecting the
0.95 quantile in SNR values and including the only 5% more differentially expressed in each subtype, we obtained 528 lncRNAs for each BRCA molecular subtype. Figure 2A shows how these lncRNAs were intersected between the five subgroups. None lncRNA was shared between all subgroups, and most of



them (over 80%) were specific for each molecular subtype. The most significant intersections were seen between LumA and LumB (44 lncRNAs), LumB and Her2 (33 lncRNAs), Basal and Normal (30 lncRNAs) and Basal and Her2 (26 lncRNAs), which represents less than 10% of the 528 lncRNAs defined for each subgroup. Even when selecting the 0.90 quantile in SNR values, none lncRNA was shared between all molecular subtypes (Supplementary Figure 2) which shows that the SNR was able to distinguish specific lncRNAs for each molecular subtype.

Secondly, we calculated the absolute SNR sum for the immune subtypes for these 528 lncRNAs selected for each BRCA molecular subgroup. Figure 2B shows the distribution of absolute SNR sum for the lncRNAs. Z-score was calculated to allow comparison between groups. The five groups demonstrated different distributions, being Basal and Her2, characterized for most lncRNAs with slight variation between the immune subtypes, while LumA and Normal presented higher variation. After selecting the 0.98 quantile (Figure 2B and Supplementary Table 3), 11 lncRNAs remained, of which only one was shared between Her2 and Basal, the lncRNA KLHDC7B-DT (ENSG00000272666). This lncRNA was removed for further analysis as we looked for a specific lncRNA signature related to each BRCA molecular subgroup. Eleven specific lncRNAs were selected for LumA, LumB and Normal and ten for Her2 and Basal. All results for the first and the second SNR calculation as well as the quantile for each lncRNA is presented in Supplementary Table 3.

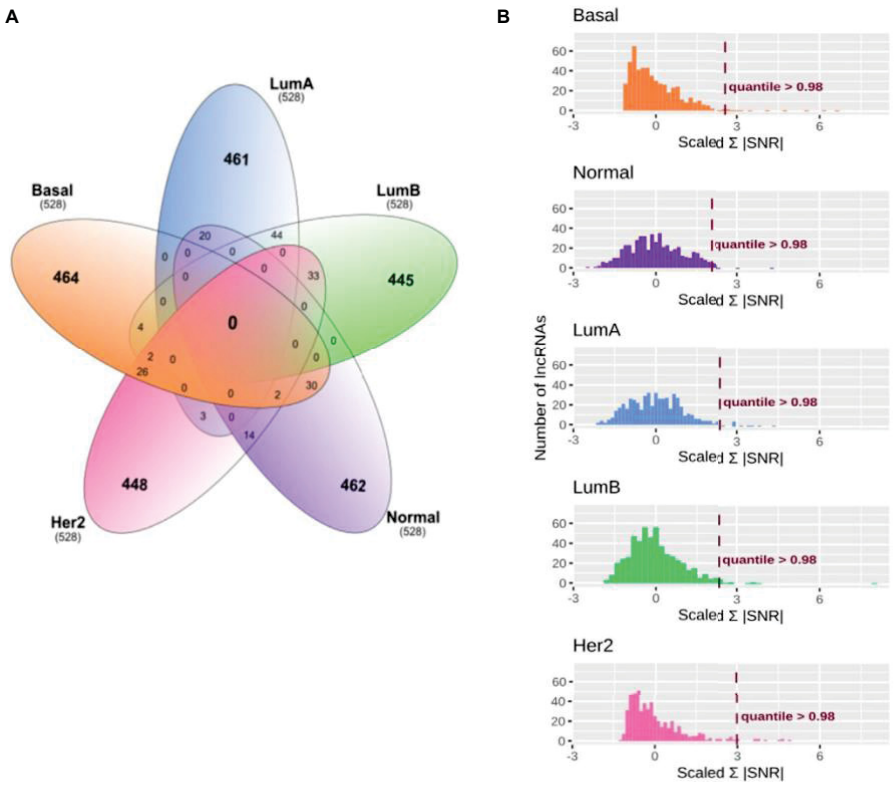


FIGURE 2 | Immune related lncRNAs filter and selection in breast cancer (BRCA) molecular subtypes. (A) Venn diagram representing specific and shared immune related lncRNAs in breast cancer molecular subtypes. After filtering for 0.95 quantile in signal to noise ratio (SNR) for the BRCA molecular subtypes, 528 lncRNAs were selected for the next analysis. (B) Absolute SNR sum filter selection. SNR was calculated within each BRCA molecular subtype for the immune subtypes. Only groups with more than five patients were used. The histograms represent the amount of lncRNAs in each range of absolute SNR sum for the BRCA molecular subtypes. X-axis was scaled for z-score to allow comparison. The lncRNAs were filtered according to the 0.98 quantile as represented as the dashed line on the histograms.

Specific lncRNAs in BRCA Molecular Subgroups Are Associated With Immune Subtypes Differentiation

The remaining specific lncRNAs related to immune subtypes are presented in Figure 3; only immune subtypes with more than five patients are shown. These were the groups used for SNR analysis. For LumA, the lncRNAs clearly distinguished the C4 and C6 subtypes while demonstrating a mixed pattern in C1, C2 and C3. Nevertheless, EWSAT1, LINC00271 and AC105285.1 show higher activation in C3 than in C2, for example. No significant correlation with OS or PFI was observed for lncRNAs expression in Cox univariate analysis in LumA patients (Figure 3A and Supplementary Table 4).

For LumB (Figure 3B), a clear distinction can be seen between C3 and C4 in the gene expression pattern, while LINC02620 and mainly AL445490.1 showed a higher expression pattern in C3. Only EBLN3P correlated with a good prognosis in both OS and PFI for Cox univariate analysis (Figure 3B and Supplementary Table 4). In the Her2 samples (Figure 3C), all ten specific lncRNAs selected demonstrated a higher expression pattern in C2 and a strong positive correlation with Leukocyte Fraction. From the Cox univariate analysis, after p-values adjustment, half of the ten lncRNAs were associated with good prognosis in PFI, but none related to OS (Figure 3C and Supplementary Table 4).

In Normal (Figure 3D), all 11 lncRNAs were suppressed in the C4 subtype and, in general, showed a higher activation pattern in C3 and less evidently in C2. Despite C6 being represented by only six patients, HLX-AS1 and AL133371.2 appeared highly expressed in the C6 group. In Cox analysis, after p-adjustment, six from the 11 lncRNAs had a hazard ratio (HR) < 1 for OS, although none presented significant values for PFI. Finally, in the Basal group (Figure 3E), nine lncRNAs presented a higher expression in C2 and a lower expression in C1, while APCDD1L-DT presented an inverse pattern. Only LINC01871 showed a significant correlation with PFI, but not for OS (Figure 3E and Supplementary Table 4).

lncRNAs Functional Annotations and Survival Analysis

We used MSigDb Hallmarks gene sets for GSEA analysis to infer possible biological roles associated with the specific lncRNAs

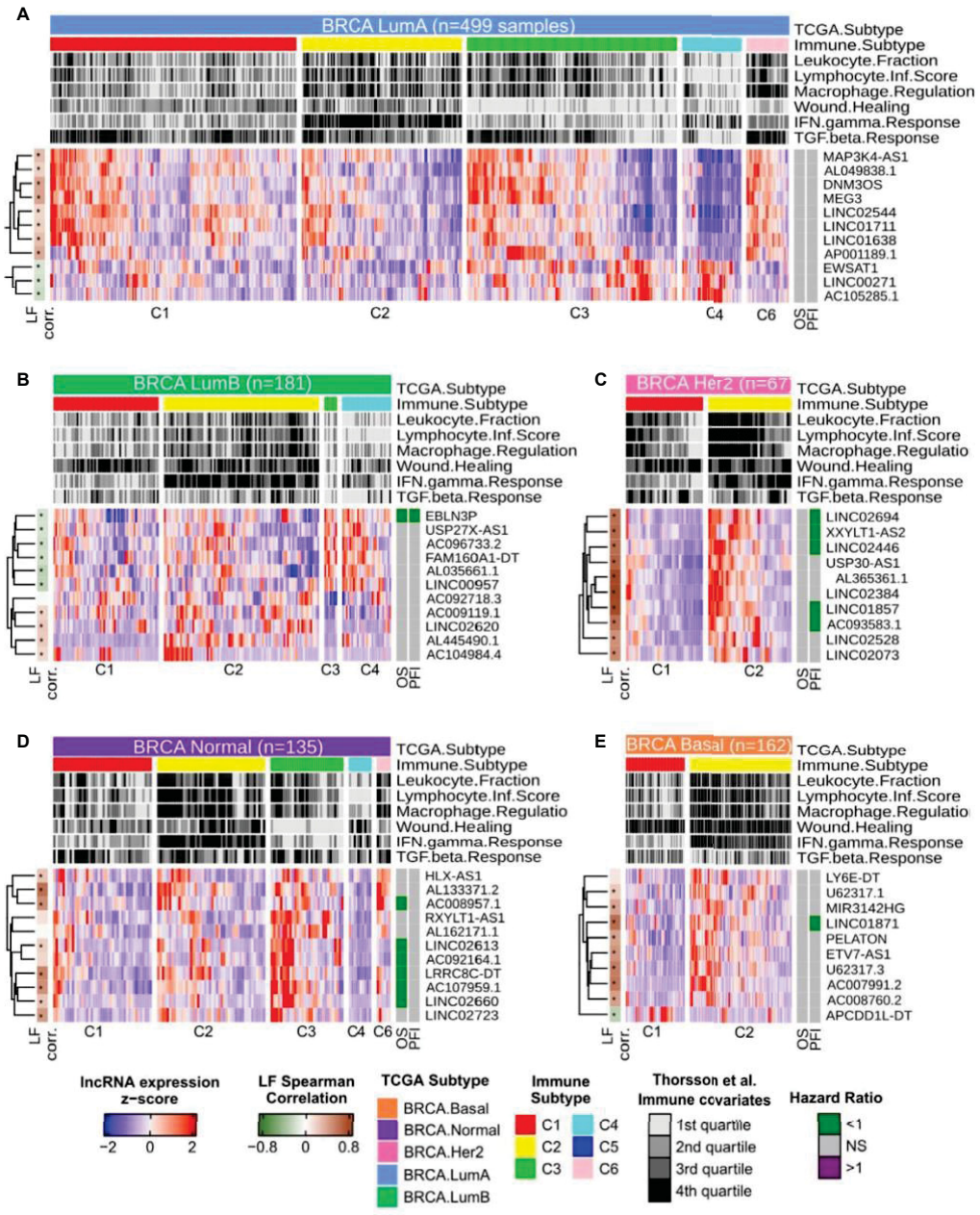


FIGURE 3 | Heatmap with column-wise z-scores for immune related lncRNAs for each BRCA molecular subtype. For color gradient, maximum and minimum z-scores were set to +2 and −2 respectively. Each column represents a sample and were semi-supervised clustered within the immune subtypes. Immune subtypes groups with less than 5 patients are not represented as they were not used in SNR calculation. The top annotations present the molecular subtype with the number of samples, the immune subtype and the immune covariates as described by Thorsson et al. (8). The immune covariates are presented as quartiles in all TCGA- BRCA primary tumor samples. The left annotation shows the Spearman's correlation for each lncRNAs with Leukocyte Fraction, being dark green for negative and
brown for positive r values, asterisks represent Spearman's correlation adjusted p-values values below 0.05. The right annotation represents the Hazard Ratio (HR) inferred by Cox Univariate analysis for Overall Survival (OS) and Progression Free Interval (PFI). Cox results with adjusted p-values greater than 0.1 are considered non significant and are colored in grey, otherwise, HR values above 1 are colored purple while HR values below 1 are colored in green. (A) LumA samples. (B) LumB samples. (C) HER2-enriched samples. (D) Normal-Like samples. (E) Basal-Like samples.

expression (Figure 3); the results for all 53 lncRNAs are presented in Supplementary Tables 5–9 and Supplementary Figure 3. We selected one lncRNA from each BRCA molecular subgroup to focus. For LumB, Her2, Basal and Normal, we chose the smallest p-value in Cox univariate analysis. As LumA did not present any significant p-values in Cox univariate analysis, we selected MEG3 as it showed a higher correlation with Leukocyte Fraction (Spearman's r=0.4, p-adjusted $< 2.2 \times 10^{-16}$).

BRCA Basal-Specific lncRNAs Are Associated With Interferon Gamma Response and Allograft Rejection Gene Sets. LINC01871 Is More Expressed in Basal and Relates to Better OS and PFI

The GSEA analysis for the selected lncRNAs from BRCA Basal group revealed immune related gene sets, like Interferon Gamma Response, Allograft Rejection and Interferon Alpha Response, enriched in all ten lncRNAs, being negatively associated only with APCDD1L-DT (Supplementary Table 5 and Supplementary Figure 3A), which was also the only one from the ten lncRNAs negatively related to Leukocyte Fraction (Supplementary Table 4). Focusing on LINC01871, this lncRNA presented a higher expression in Basal than all other BRCA molecular subtypes (Figure 4A) and was mainly associated with immune activation, for instance, enriched in Interferon Gamma Response, Allograft Rejection, Interferon Alpha Response and Inflammatory Response gene sets (Figure 4D). Within Basal samples, it appears significantly suppressed in C1 (Figure 4B), while in

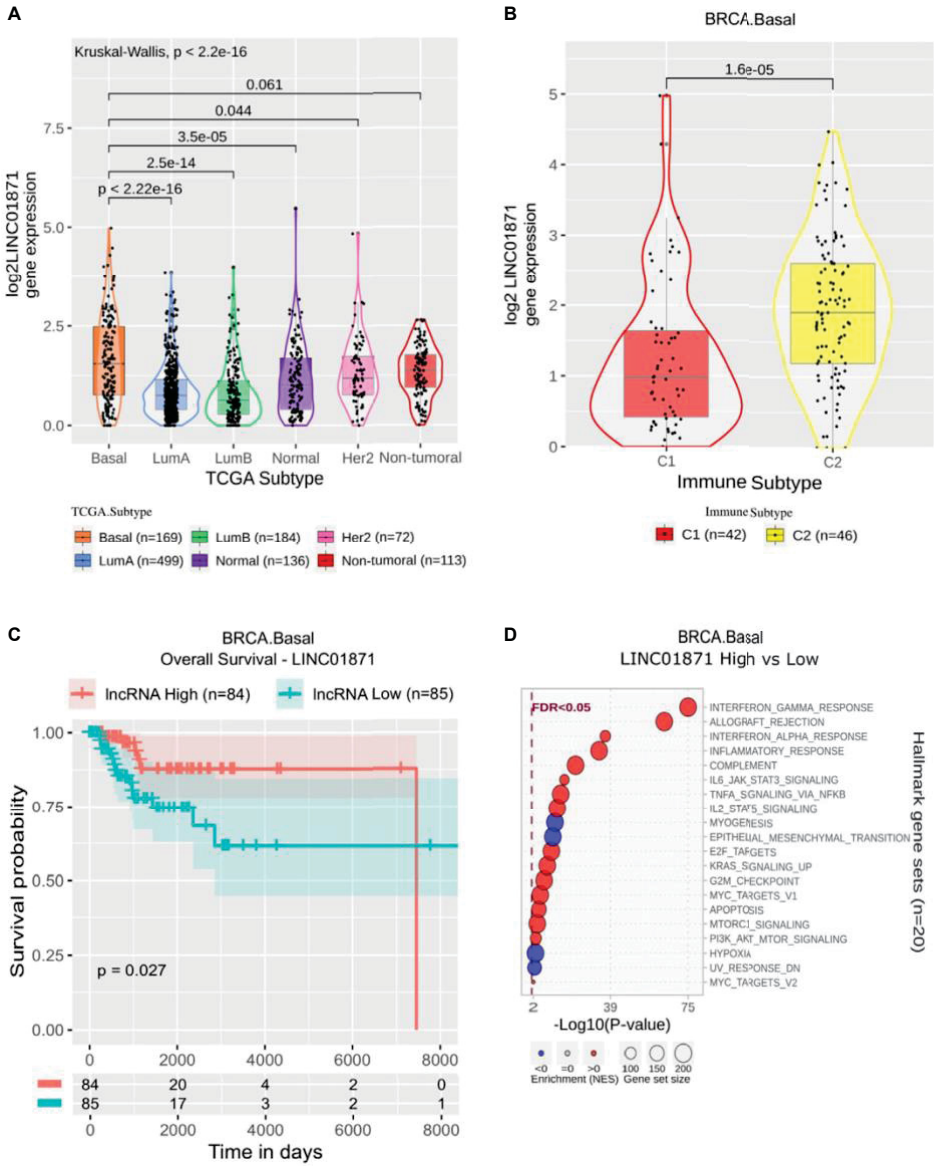


FIGURE 4 | LINC01871 panel representation in basal-like subtype. (A) Box-plots representing LINC01871 expression among all breast cancer molecular subtypes. The Kruskal-Wallis test (p-value represented in the panel left top position) was used to differentiate expression between groups, followed by the Wilcoxon test (p-values represented emphasizing comparisons between subtypes). (B) Box-plot representing LINC01871 in C1 and C2 immune subtypes in basal-like patients. (C) LINC01871 overall survival curves. To define the two groups, LINC01871's median expression classified the patients into "high" and "low" groups. Logrank p-value represented as 0.027. (D) Enrichment analysis using Hallmarks gene sets. Red circles refers to activation while blue circles to inactivation. In the X axis, a p-value scale is represented. The circle size varies according to the number of genes in the identified gene set.

Kaplan-Meier analysis, it was associated with better OS and PFI (Figure 4C and Supplementary Figure 4A).

BRCA Normal Specific lncRNAs Are Associated With Suppression of Proliferation Gene Sets. LINC02613 Is Suppressed in C4 and Relates to OS

The 11 BRCA Normal specific lncRNAs were related to the suppression of gene sets like G2M checkpoint, E2F and MYC targets associated with proliferation (Supplementary Table 6 and Supplementary Figure 3B). Mainly AL133371.2, AC008957.1, LRRC8C-DT, AC107959.1, LINC02660 and LINC02723 presented immune-related gene sets positively enriched (Allograft Rejection, Coagulation, Complement, IL6 JAK STAT3 Signaling, Inflammatory, Interferon Alpha and Interferon Gamma Response). LINC02613 follows this enrichment pattern (Figure 5D). It is worth noting that Estrogen Response Early and Late gene sets appear suppressed only for this lncRNA and AC092164.1. Although its expression significantly differs from BRCA Normal to all other molecular subtypes (Figure 5A), it is more expressed in non-tumoral samples followed by BRCA Basal and Normal. At the same time, it seems suppressed in LumA, LumB and Her2.

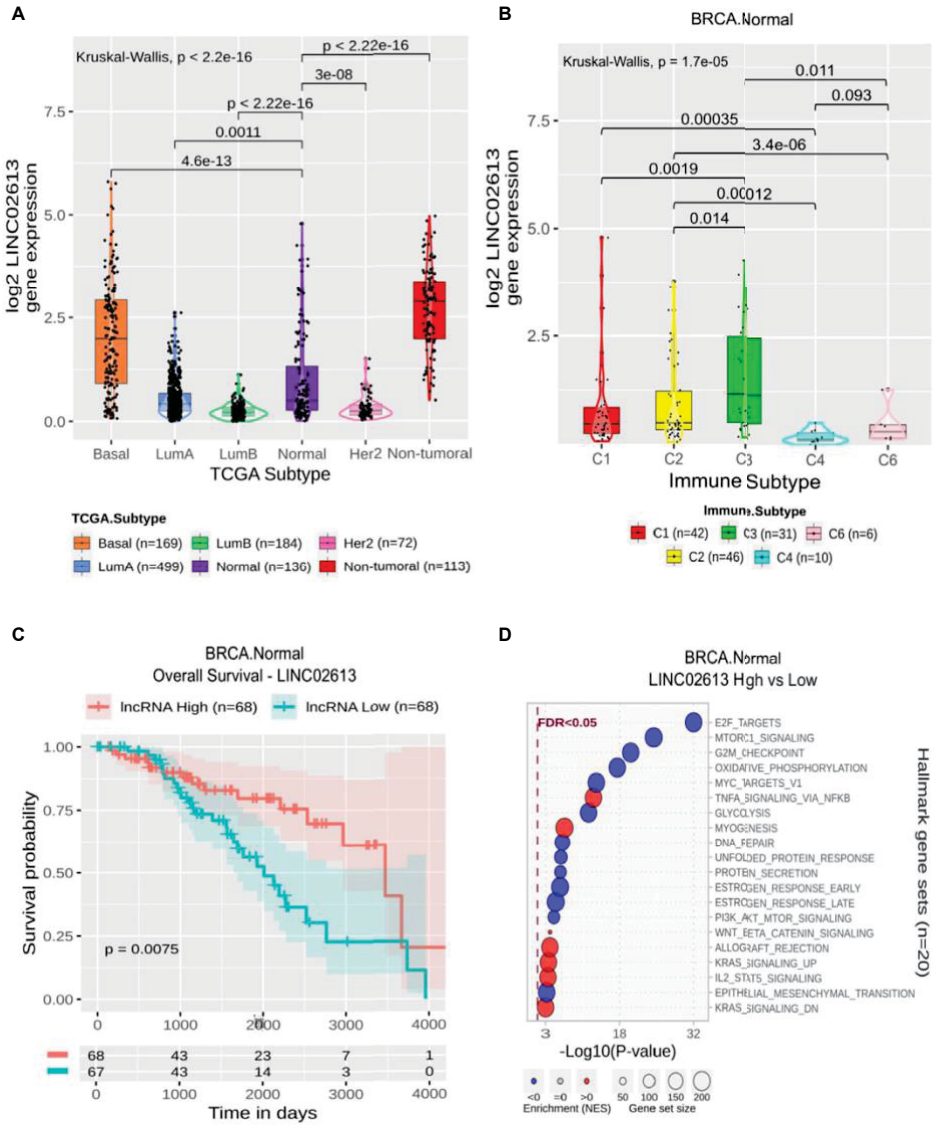


FIGURE 5 | LINC02613 panel representation in normal-like subtype. (A) Box-plots representing LINC02613 expression among all breast cancer molecular subtypes. The Kruskal-Wallis test (p-value represented in the panel left top position) was used to differentiate expression between groups, followed by the Wilcoxon test (p-values represented emphasizing comparisons between subtypes). (B) Box-plot representing LINC02613 in C1,C2, C3, C4 and C6 immune subtypes in normal-like patients. (C) LINC02613 overall survival curves. To define the two groups, LINC02613's median expression classified the patients into "high" and "low" groups. Logrank p-value represented as 0.0075. (D) Enrichment analysis using Hallmarks gene sets. Red circles refers to activation while blue circles to inactivation. In the X axis, a p-value scale is represented. The circle size varies according to the number genes in of the identified gene set.

Within Normal samples, it is more expressed in C3 and significantly suppressed in C4 (Figure 5B) and correlates with a better OS (Figure 5C) but not in PFI (Supplementary Figure 4B).

LumA-Specific lncRNAs Relate to Epithelial-Mesenchymal Transition, Immune and Proliferation Gene Sets. MEG3 Is Suppressed in BRCA Compared With Non-Tumoral Samples

For LumA (Supplementary Table 7 and Supplementary Figure 3C), the Epithelial-Mesenchymal Transition (EMT) module was enriched in all 11 lncRNAs, being EWSAT1, LINC00271, and AC105285.1 associated negatively with this module and the other eight lncRNAs positively. In general, this pattern was followed for the immune-related gene sets and the correlation with Leukocyte Fraction (Figure 3 and Supplementary Table 4), that is to say, EWSAT1, LINC00271, and AC105285.1 with the negative association and the other lncRNAs with a positive one. Gene sets associated with proliferation (E2F targets, Myc targets V1 and V2) were overall suppressed in all lncRNAs. Oxidative phosphorylation appeared suppressed in the lncRNAs positively associated with immune gene sets (MAP3K4-AS1, AL049838.1, DNM3OS, MEG3, LINC02544, LINC01711, LINC01638, and AP001189.1).

MEG3 presented a lower expression on BRCA compared with non-tumoral samples and a significant difference when comparing its expression in LumA with other BRCA molecular subtypes except Normal (Figure 6A). Within LumA samples, the immune subtypes comparisons resulted in a p-value of 3.2×10^{-14} in the Kruskal-Wallis test with a higher expression pattern in C3 and C6. At the same time, C4 stands out with a significantly lower pattern compared with all other immune subtypes (Figure 6B). In general, the GSEA analysis resulted in modules associated with immune response enriched with MEG3 overexpression and modules related to proliferation enriched with MEG3 suppression (Figure 6D). As expected by the Cox analysis, the Kaplan-Meier did not significantly impact MEG3 expression on OS or PFI (Figure 6C and Supplementary Figure 4C).

Most Specific lncRNAs in LumB Are Associated With Immune Pathways. EBLN3P Is Associated With a Good Prognosis in OS and PFI

LumB specific lncRNAs demonstrated a higher variation in the enriched modules (Supplementary Table 8 and Supplementary Figure 3D). AC009119.1, LINC02620, AL445490.1, and AC104984.4 high expressions were positively related with immune gene sets, while with exception of LINC02620, no clear relation was observed for the proliferation ones. These four lncRNAs were also positively correlated with Leukocyte Fraction (Supplementary Table 4). AC092718.3, which did not present a significant correlation with Leukocyte Fraction, was also not related to the enrichment of immune gene sets; otherwise, its enrichment revealed proliferation modules upregulated. USP27X-AS1, AC096733.2, FAM160A1-DT, AL035661.1 and LINC00957 were associated with suppressing gene sets like Allograft Rejection, Interferon Gamma Response and Inflammatory Response; they also related negatively with Leukocyte Fraction (Supplementary Table 4).

EBLN3P is also related negatively to Leukocyte Fraction and with some immune modules. Still, its enrichment revealed a mixed pattern of modules suppressed like Epithelial-Mesenchymal Transition, Oxidative Phosphorylation, Myogenes and Myc Targets (Figure 7D). In LumB, its expression was higher in C3 and C4 than C1 and C2 (Figure 7B). In comparison, it presented a significantly higher expression in LumB samples than in the other BRCA molecular subtypes except for LumA (Figure 7A). In LumB samples, the higher expression of EBLN3P was related to a better outcome in OS but not with PFI in Kaplan-Meier analysis (Figure 7C and Supplementary Figure 4D).

Her2-Specific lncRNAs Are Associated With Immune Activation and XXYLT1-AS2 Associated With Higher PFI

All 10 Her2 specific lncRNAs were related with immune activation as can be seen in their enrichment pattern of Allograft Rejection, Interferon Gamma and Alpha Response, Inflammatory Response, IL6 JAK STAT3 signaling and Complement (Supplementary Table 9 and Supplementary Figure 3E) and in the strong positive correlation with Leukocyte Fraction (r range from 0.49 to 0.82 in Spearman's Correlation, p-adjusted from 1.5×10^{-5} to $<2.2 \times 10^{-16}$). LINC02384 and LINC02073 are also related to the suppression of proliferation pathways (E2F Targets, G2M checkpoint and MYC Targets).

XXYLT1-AS2 showed a significantly higher expression in Her2 than in LumA, LumB and Normal (Figure 8A). Within Her2 samples, it was significantly more expressed in the C2 immune subtype (Figure 8B) and associated with better prognosis in PFI (Figure 8C) but not with OS (Supplementary Figure 4E). Its enrichment revealed high positive correlation with Allograft Rejection, Interferon Gamma Response, Inflammatory Response and IL6 and IL2 signaling and negative relation mainly with EMT, Hypoxia and Myogenesis (Figure 8D).

DISCUSSION

Using breast cancer immunogenomics data already published (8), we propose here five distinct immune-related lncRNAs signature according to BRCA molecular subtypes using a two steps SNR selection. Each molecular subtype presented a specific immune-related lncRNAs signature and in GSEA, in general, these lncRNAs functions varied between proliferation and immune activation or suppression, which demonstrates that our selection methodology was able to filter lncRNAs related to the immune response. The survival impact of the selected lncRNAs diverged across the molecular subtypes, in agreement with the fact that the immune activation also differs in terms of prognosis importance between the molecular subtypes (24). For instance, the tumor-infiltrating lymphocytes (TIL) was not related to prognosis in Estrogen Positive BRCA tumors, which

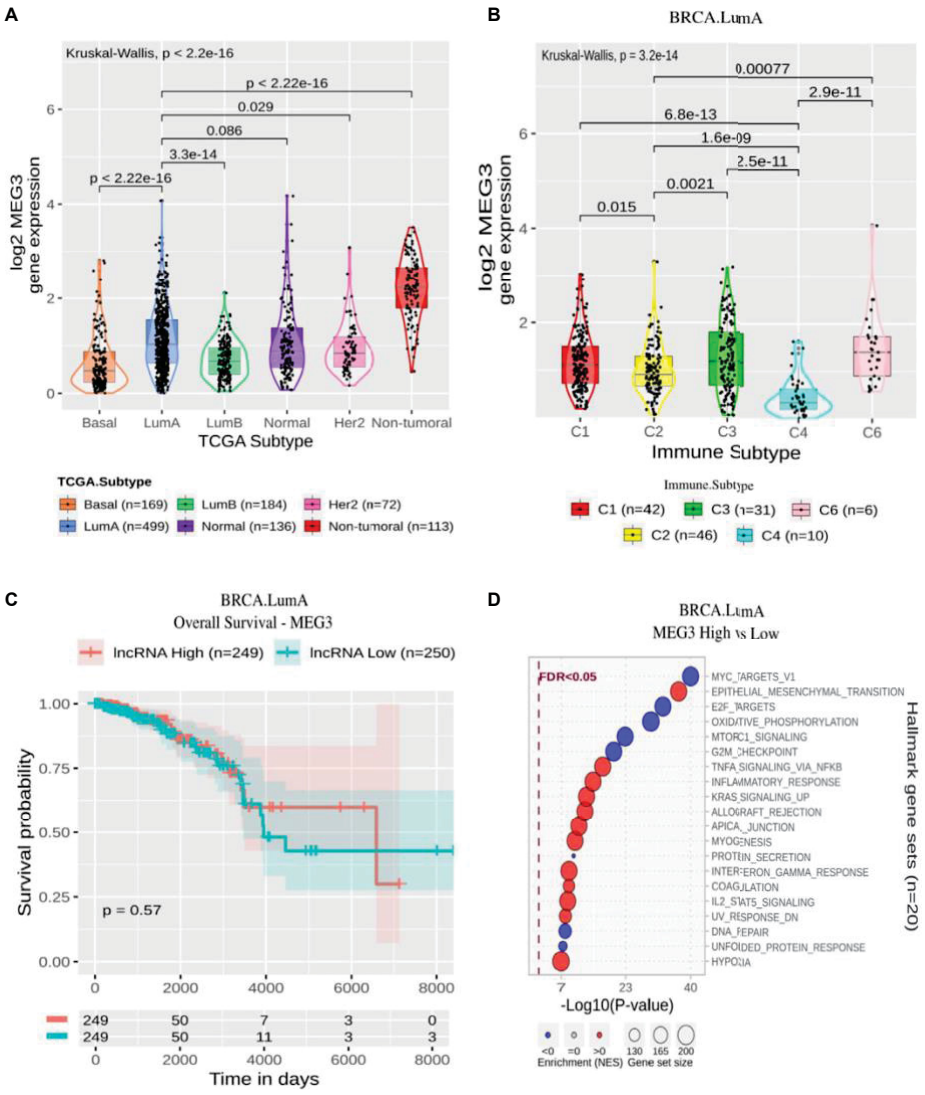


FIGURE 6 | MEG3 panel representation in LumA subtype. (A) Box-plots representing MEG3 expression among all breast cancer molecular subtypes. The Kruskal- Wallis test (p-value represented in the panel left top position) was used to differentiate expression between groups, followed by the Wilcoxon test (p-values represented emphasizing comparisons between subtypes). (B) Box-plot representing MEG3 in C1,C2, C3, C4, and C6 immune subtypes in LumA patients.
(C) MEG3 overall survival curves. To define the two groups, MEG3's median expression classified the patients into "high" and "low" groups. Logrank p-value represented as 0.57. (D) Enrichment analysis using Hallmarks gene sets. Red circles refers to activation while blue circles to inactivation. In X axis, a p-value

may explain the lack of immune-related lncRNAs associated with OS or PFI in LumA/B (24). Nevertheless, most lncRNAs selected for Her2 and Normal were associated with PFI and OS, respectively, reinforcing BRCA molecular subtypes' well-known heterogeneity. In the article used as a reference (8), the authors elaborated a list of 75 immunomodulatory genes. With this in mind, in Supplementary Figure 1, we represented the expression variation (as log2 gene expression z-score) among the molecular subtypes, also considering the immune-related subtype. It is possible to observe a distinct expression pattern in all molecular groups, as expected, regarding breast cancer as a heterogeneous disease. This result reflects the high heterogeneity observed in BRCA molecular subtypes, emphasizing the relevance of characterizing them better molecularly, and we included lncRNA analysis in this complexity.

A specific immune-related signature was proposed to LumA subtype using 11 lncRNAs. Among them, three were suggested related to immune response repression (AC105285.1, LINC00271 and EWSAT1), and none of these were previously investigated under BRCA or immune response aspects (Supplementary Table 7 and Supplementary Figure 3C). However, LINC00271 and EWSAT1 had already been

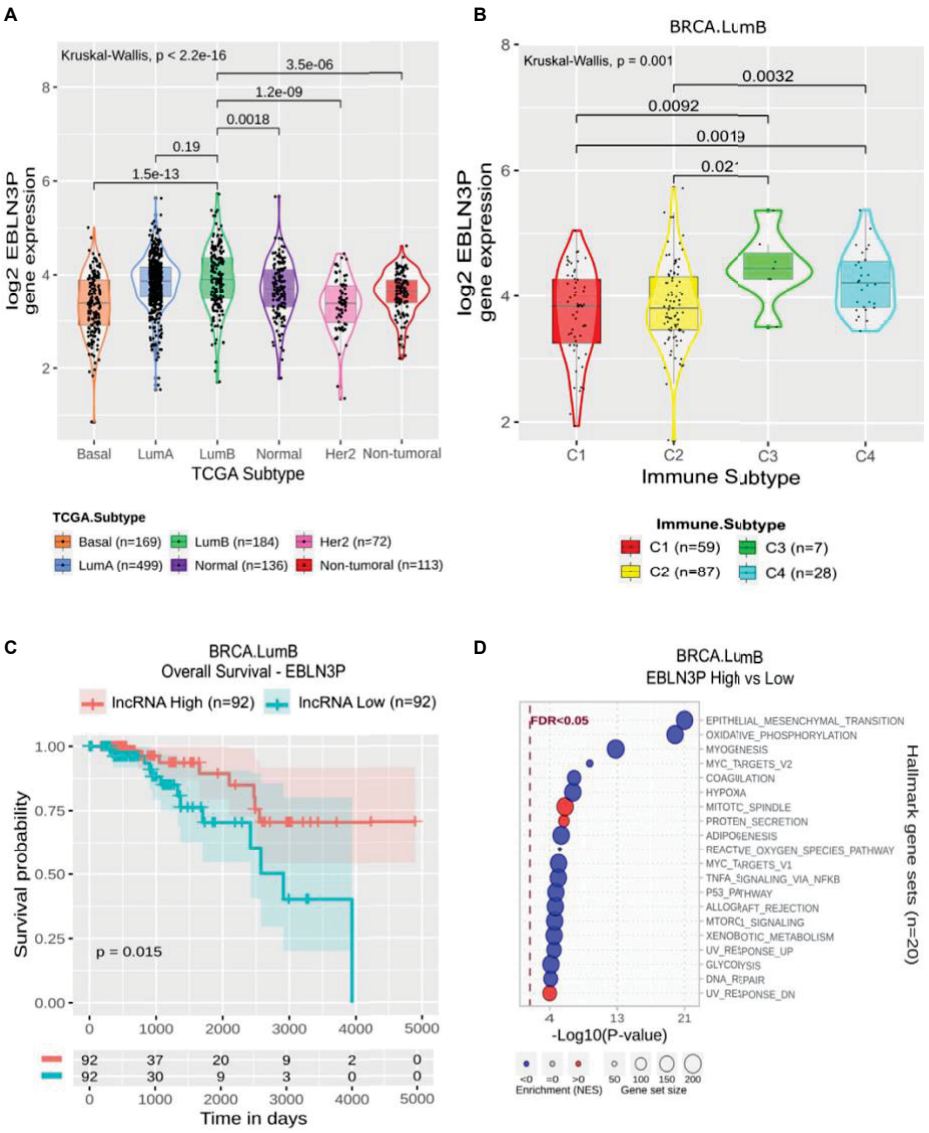


FIGURE 7 | EBLN3P panel representation in LumB subtype. (A) Box-plots representing EBLN3P expression among all breast cancer molecular subtypes. The Kruskal-Wallis test (p-value represented in the panel left top position) was used to differentiate expression between groups, followed by the Wilcoxon test (p-values represented emphasizing comparisons between subtypes). (B) Box-plot representing EBLN3P in C1,C2, C3, and C4 immune subtypes in LumB patients.
(C) EBLN3P overall survival curves. To define the two groups, EBLN3P's median expression classified the patients into "high" and "low" groups. Logrank p-value represented as 0.015. (D) Enrichment analysis using Hallmarks gene sets. Red circles refers to activation while blue circles to inactivation. In X axis, a p-value scale is
represented. The circle size varies according to the number of genes in the identified gene set.

associated with other cancer types. LINC00271 low expression was associated with poor prognosis in papillary thyroid cancer (25) and in adrenocortical tumors (26). EWSAT1 was associated with progression in several cancer types, such as ovarian (27), cervical (28) and colorectal (29).

The lncRNA maternally expressed gene 3 (MEG3) is highlighted here in LumA immune response context. MEG3 is up-regulated in C3 and C6 subtypes and is related to neither overall nor progression aspects (Figure 6). The immune subtypes C3 and C6 demonstrate high scores of lymphocyte infiltrate,

macrophage regulation and TGF-b response (Figure 3). Thus, the increased expression of this lncRNA can be related to immune response activation. Indeed, our functional characterization revealed that MEG3 high expression is associated with several immune hallmarks, such as: "TNFA Signaling Via NFKB," "Inflammatory Response," and "Interferon Gamma Response" (Figure 6D). MEG3's role in BRCA immune response is to the best of our knowledge the first time cited here. This lncRNA has only been described in endometrial cancer cells down-regulating *PD-L1* (30). MEG3 is found downregulated in several cancer types,

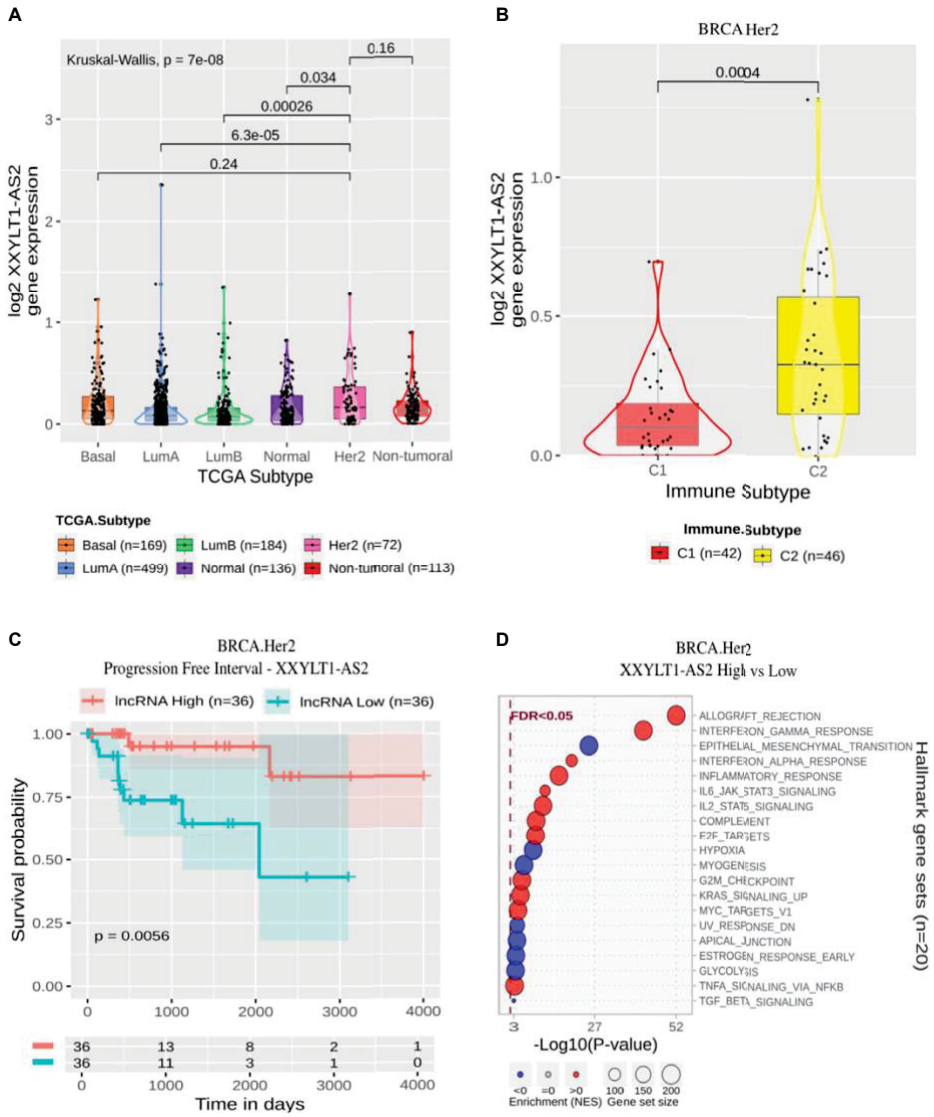


FIGURE 8 | XXYLT1-AS2 panel representation in Her2 subtype. (A) Box-plots representing XXYLT1-AS2 expression among all breast cancer molecular subtypes. The Kruskal-Wallis test (p-value represented in the panel left top position) was used to differentiate expression between groups, followed by the Wilcoxon test (p- values represented emphasizing comparisons between subtypes). (B) Box-plot representing XXYLT1-AS2 in C1, C2, C3, and C4 immune subtypes in Her2 patients.
(C) XXYLT1-AS2 progression free interval curves. To define the two groups, XXYLT1-AS2's median expression classified the patients into "high" and "low" groups. Logrank p-value represented as 0.0056. (D) Enrichment analysis using Hallmarks gene sets. Red circles refer to activation while blue circles to inactivation. In X axis, a p-value scale is represented. The circle size varies according to the number of genes in the identified gene set.

such as BRCA, liver, colorectal and cervical cancer and was experimentally evidenced as *TP53*'*s* regulator (31). In BRCA, MEG3's downregulation is associated with poor overall survival and tumor staging (32).

In LumB subtype, 11 specific lncRNA were selected, and five of them (LINC00957, AL035661.1, FAM160A1-DT, AC096733.2 and
EBLN3P) have their expression related to immune response hallmarks repression (Supplementary Table 8 and Supplementary Figure 3D). These lncRNAs were detected up-regulated in immune subtypes with lower lymphocyte fraction, such as C4. Among these lncRNAs only LINC00957, AL035661.1, and EBLN3P had already been studied in cancer context. LINC00957 high expression was associated with bad prognosis in colorectal cancer (33) and osteosarcoma (34). The lncRNA AL035661.1 was found in a lncRNA profile that managed to efficiently predict early recurrence in hepatocellular carcinoma after curative resection (35) and was associated with EMT in Kidney renal clear cell carcinoma (36).

The lncRNA endogenous Bornavirus-like nucleoprotein 3 (EBLN3P) was highlighted in the LumB subtype. EBLN3P is still not well characterized in the literature since few published

studies have focused on its mechanisms and effects on human diseases (37). Dai and colleagues (38) noted that EBLN3P is expressed in osteosarcoma tissues and cell lines. They pointed out that its overexpression promotes proliferation, migration and invasion. Li et al. (37) have already reported EBLN3P as a novel oncogene for liver cancer for similar aspects (37, 38).

EBLN3P's expression is higher in C3 (Inflammatory) and C4 (Lymphocyte depleted) immune subtypes in LumB samples (Figure 7B). In fact, in LumB samples, these immune subtypes exhibit a low TGF-b response score. The role of TGF-b response is still controversial in cancer, depending on the tumor stage. Indeed, TGF-b promotes EMT (39), thus being related to cell proliferation. According to the immune subtypes described by Thorsson et al. (8), the C3 subtype exhibits a low proliferation rate and C4 a moderate one. Using this data with our analysis, we observed a concordance of the data since EBLN3P negatively correlates with TGF-b (Supplementary Table 8). In this way, while this lncRNA is highly expressed in these subtypes, TGF-b has a lower expression. The lncRNA EBLN3P's expression was associated with better survival and disease progression outcomes (Figure 7C). Regarding its enrichment analysis in LumB samples, the most significant result was achieved considering "Epithelial Mesenchymal Transition" gene sets. In this scenario, EBLN3P's low expression is related to EMT activation.

Interestingly, most LumA and LumB lncRNAs were related positively or negatively to the Interferon Gamma Response module. Recently, a study demonstrated that the activation of the interferon signaling pathway in ER+ BRCA relates to resistance to CDK4/6 inhibitors and immune checkpoint activation (40). Understanding the role of lncRNAs in this context may be the subject of future studies.

A total of 10 Basal lncRNAs were selected after SNR analysis. According to functional annotation analysis of these lncRNAs. (Supplementary Table 5 and Supplementary Figure 3A), only the lncRNA APCDD1L-DT's high expression was related to immune response inhibition. This lncRNA was only evaluated in a competing endogenous RNA network in lung cancer (41), so its immune response regulation role needs to be better understood. The remaining nine lnRNAs were associated with immune response activation (Supplementary Figure 3A), and only MIR3142HG was previously studied in cancer context. MIR3142HG polymorphisms were associated with glioma susceptibility and/or prognosis (42). An interesting fact is that MIR3142HG was evidenced as a regulator of IL-1b induced inflammatory response in lung fibroblasts (43). In our analysis, inflammatory response appears to be induced by MIR3142HG (Supplementary Figure 3A).

In the Basal, we highlight the lncRNA LINC01871. According to immune subtypes classification, LINC01871 is up-regulated in C2 samples, which is called IFN-g dominant, and its high expression is associated with a better prognosis (Figure 4). Our enrichment analysis (Figure 4D) showed a strong relation to immune processes. Considering only Basal samples, LINC01871 was associated with "Allograft rejection," "Interferon Gamma," "Interferon Alpha," and "Inflammatory," for example. In these processes, high expression LINC01871 was associated with the activation of these immune responses. We observe a strong relationship with the leucocyte fraction, which might be related to this observed immune response activation. These results emphasize the relevance of LINC01871 in immune response activation in breast tumor samples, especially in Basal samples, which might be related to better prognosis response.

This lncRNA has already been related to immunity in breast, cervical and gastric cancer. LINC01871 was detected in an immune-related prognosis signature in BRCA and exhibited a strong positive correlation with genes associated with immune response such as *GZMB*, *CTLA4* and *PDCD1*. The authors suggested that this lncRNA may play an important role, mainly related to the above immune processes and immune genes (44). In cervical cancer, LINC01871 was also found in an immune-related prognostic signature being related to immune response and TGF-b signaling pathway (45); additionally, in gastric cancer, LINC01871 expression was positively correlated with CD8+ T cell enrichment levels, cytolytic immune activity, and *CD274* (*PD-L1*) expression levels in TCGA gastric cancer cohort (46).

Taking together our results and background literature is possible to correlate LINC01871 with a cytotoxic immune response. Our enrichment analysis showed an association with Interferon alpha and gamma, which are strongly related to cytotoxic response, going in line with what was observed in other studies (44, 46). In BRCA, CD8+ T cells are reported as prognostic significance in estrogen receptor (ER)-negative BRCA, but not in ER-positive cases, being associated with better clinical outcomes survival and response to treatment (47). Also, CD8+ responses have significantly elevated expression of multiple immune checkpoint molecules, such as programmed cell death 1 (PD-1), programmed death-ligand 1 (PD-L1) and 2 (PD-L2) and cytotoxic T-lymphocyte-associated protein 4 (CTLA4) (48). In this way, a stronger infiltration of CD8+ T cells can predict patient response to standard of care chemotherapy and immune checkpoint blockade therapy, such as anti-PD-1 (49).

According to Figure 3, it is possible to observe a distinct pattern of expression of LINC01871, both considering immune subtype and Interferon Gamma response. Recently, Interferon Gamma low expression level was associated with worse prognosis in Basal patients (50). In our analysis, LINC01871 high expression was associated with better prognosis and Interferon Gamma response activation (Figure 4D), reinforcing the fact that LINC01871 can be used as a good prognosis marker for Basal patients. This can also be evaluated focusing on therapy response, mainly on immunotherapy. Recently, a combination of an immunotherapy drug with chemotherapy was approved for metastatic triple-negative BRCA (51); however, due to this group's high heterogeneity, only a fraction of patients respond well to this treatment. Yang and colleagues (52) defend that an immunity score may be used together with PD-L1 expression to a better design for trials testing immune-checkpoint inhibitors. In this context, lncRNAs like LINC01871 may be used to enhance this selection criterion. However, to be fully applicable as a biomarker, the molecular aspects by which LINC01871 is involved in the immune system activation process need to be better understood.

In the Her2 subtype, ten lncRNAs were evidenced using our selection methodology. All of these lncRNAs were related to immune response activation, as represented in Supplementary Table 9 and Supplementary Figure 3E. Among the ten lncRNAs, four of them had already been related to cancer. LINC02446 is associated with prognosis (53) and EMT activation (54) in bladder cancer. Similarly, USP30-AS1 is associated with prognosis in bladder cancer (54). This lncRNA was also associated with autophagy in ovarian cancer (55), and was related to immune response in cervical cancer (56) and glioblastoma (57). AL365361.1 was associated with good prognosis and immune response in head and neck squamous cell carcinoma (58) and with early recurrence in hepatocellular carcinoma (35). The lncRNA LINC01857 was related to progression in gastric cancer (59), glioma (60), BRCA (61) and B-cell Lymphoma (62). Thus, this lncRNA can be suggested to act as an oncogene. We highlight the lncRNA XXYLT1-AS2, which was associated with progression-free survival in Cox proportional hazard ratio analysis (HR = 0.0011; p-adjusted = 4×10^{-2}). XXYLT1-AS2 is up-regulated in C2 subtype (Figure 8B), and its high expression is related to better progression-free interval (p- logrank = 5.6×10^{-3}). According to our functional annotation analysis, XXYLT1-AS2 is associated with immune response activation and EMT repression (Figure 8D). This result converges with what we observed concerning the progression- free interval since EMT is one of the main pathways activated during the disease progression process. XXYLT1-AS2 is, to the best of our knowledge, the first time described associated with cancer. This lncRNA was only evaluated in Human umbilical vein endothelial cells (HUVECs), and its up-regulation was related to the inhibition of these cell's proliferation and migration (63).

Finally, for the Normal subtype, a signature composed of 11 lncRNAs was predicted. Most of them are associated with immune response activation and overall survival, and we discuss here LINC02613. This lncRNA is up-regulated in the C3 subtype (Figure 5B) and is significantly related to the patient's overall survival. LINC02613's high expression is associated with a better prognosis (p-logrank = 7.5×10^{-3}). The enrichment analysis evidenced that LINC02613 is related to immune response activation and mainly involves cell cycle repression (Figure 5D). Indeed, the C3 subtype was described as one with the lowest proliferation rates (8). LINC02613 has not been described in any biological context yet. So we emphasize here the relevance to better characterize this lncRNA.

Considering all the generated and discussed data in this work, we highlight the applied methodology's significance to look for immune-related lncRNAs. The filtering steps followed by functional annotation efficiently got lncRNAs that might be related to immune response. The role of lncRNAs in regulating immune response is increasingly being explored. Our study is limited to *in silico* analysis; however it brings up new lncRNAs candidates as it is a hypothesis generator article.

Another aspect that is important to discuss here is that we found distinct lncRNAs signature for each molecular subtype that may help find important lncRNAs in the immune response process that may be used to guide therapy candidates or as biomarkers; also our results point out that different lncRNAs may be implicated in immune response depending on BRCA molecular subtypes. Our findings are in agreement with what has been discussed about the heterogeneity of BRCA. We identified that different lncRNAs in each molecular subtype are related to the immune system activation. For example, MEG3 and LINC01871 were associated with activation of Interferon Gamma, in LumA and Basal, respectively. This finding highlights the importance of molecularly characterizing each subtype in order to enable increasingly personalized therapeutic approaches.

Nonetheless, the lncRNAs presented here certainly do not cover all important immune-related lncRNAs in BRCA. Our focus was to find lncRNAs that, in some order, could play a significant role in the immune distinction for each molecular subgroup.

CONCLUSION

In conclusion, we present a BRCA specific molecular subtype immune-related lncRNAs signature that may guide future studies aiming to look for important biomarkers in BRCA and highlight the relevance of lncRNAs in the immune subtype's classification.

DATA AVAILABILITY STATEMENT

All scripts, datasets, software and algorithms used to generate results, figures, and tables for this study are available on the GitHub repository (https://github.com/sysbiolab/Sup_Material_Mathias2021) and Supplementary Table 10.

AUTHOR CONTRIBUTIONS

CM, JM, and BA conceived the presented idea. CM, JM, and BA developed the theory and performed the computations. DG, MC, and JO verified the analytical methods. DG, MC, and JO supervised the findings of this work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Public Research Agencies CAPES (001-Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) processo 153771/2018-6. JM received scholarship from the BIG DATA innovation program from the Associação Hospitalar de Proteção à Infância Raul Carneiro- AHPIRAC (2020).

ACKNOWLEDGMENTS

Data used in this manuscript were obtained from The Cancer Genome Atlas (TCGA) BRCA database.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.692170/full#supplementary-material

Supplementary Figure 1 | Heatmap of Immune modulators gene set (n=74) proposed by the reference article Thorsson et al. (8) in breast cancer molecular subtypes. For color gradient, maximum and minimum column-wise z-scores were set to +2 and −2 respectively. Each column represents a sample and were semi-
supervised clustered within the molecular subtypes.

Supplementary Figure 2 | Venn diagram representing specific and shared immune related lncRNAs in breast cancer (BRCA) molecular subtypes after filtering for 0.90 quantile in signal to noise ratio (SNR) for the BRCA molecular subtypes. 593 lncRNAs were filtered by this criterion for each molecular subtype.

Supplementary Figure 3 | Enrichment map for immune-related lncRNA signature for each molecular subtype: (A) Basal, (B) LumA, (C) LumB, (D) Normal, and (E) Her2. The lncRNAs are organized in the X axis and Hallmarks gene sets in the Y axis. Red circles refer to positive enrichment scores while blue circles to negative. The circle size varies according to -log10 (adjusted p-value), bigger circles mean greater significance.

Supplementary Figure 4 | Kaplan-Meier analysis for overall survival for LINC01871 in Basal (A) and XXYLT1-AS2 in Her2 (E) and for Progression Free Interval for MEG3 in LumA (B), EBLN3P in LumB (C) and LINC02613 in Normal (D). The groups of High and Low expressions are based on the median value of its respective lncRNA expression. Confidence interval and logrank p-value are shown
and values below 0.05 are considered significant.

Supplementary Table 1 | Breast cancer cohort stratification into the immune subtypes.

Supplementary Table 2 | Ensembl and HGNC Symbol from lncRNAs used in this study.

Supplementary Table 3 | First and second SNR calculated for lncRNAs for each molecular subtype.

Supplementary Table 4 | Cox Proportional Hazard models for the specific immune-related lncRNAs and Leukocyte correlation.

Supplementary Tables 5–9 | GSEA Hallmarks analysis for the specific immune-related lncRNAs organized by the molecular subtype.

Supplementary Tables 10 | Data sets, software and algorithms used in this study.

REFERENCES

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular Portraits of Human Breast Tumours. *Nature* (2000) 406:747–52. doi: 10.1038/35021093
2. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz J, et al. Phenotypic and Molecular Characterization of the Claudin-Low Intrinsic Subtype of Breast Cancer. *Breast Cancer Res* (2010) 12(5):R68. doi: 10.1186/BRCAr2635
3. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene Expression Patterns of Breast Carcinomas Distinguish Tumor suBRCAlasses With Clinical Implications. *Proc Natl Acad Sci USA* (2001) 98(19):10869–74. doi: 10.1073/pnas.191367098
4. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast Cancer Intrinsic Subtype Classification, Clinical Use and Future Trends. *Am J Cancer Res* (2015) 5 (10):2929–43. doi: 10.1371/journal.pone.0124964
5. Shien T, Iwata H. Adjuvant and Neoadjuvant Therapy for Breast Cancer. *Jpn J Clin Oncol* (2020) 50(3):225–9. doi: 10.1093/jjco/hyz213
6. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol* (2020) 27(1):27–35. doi: 10.1097/PAP.0000000000000232
7. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive Evaluation of Transcriptome-Based Cell-Type Quantification Methods for Immuno-Oncology. *Bioinformatics* (2019) 35:i436–45. doi: 10.1093/bioinformatics/btz363
8. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023
9. Segovia-Mendoza M, Morales-Montor J. Immune Tumor Microenvironment in Breast Cancer and the Participation of Estrogen and Its Receptors in Cancer Physiopathology. *Front Immunol* (2019) 10:348. doi: 10.3389/fimmu.2019.00348
10. Hombach S, Kretz M. Non-Coding RNAs: Classification, Biology and Functioning.
Adv Exp Med Biol (2016) 937:3–17. doi: 10.1007/978-3-319-42059-2_1
11. Clark MB, Mercer TR, Bussotti G, Leonardi T, Haynes KR, Crawford J, et al. Quantitative Gene Profiling of Long Noncoding RNAs With Targeted RNA Sequencing. *Nat Methods* (2015) 12(4):339. doi: 10.1038/nmeth.3321
12. Chan JJ, Tay Y. Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci* (2018) 19(5):1310. doi: 10.3390/ijms19051310
13. Bolha L, Ravnik-GlavacM, GlavacD. Long Noncoding RNAs as Biomarkers
in Cancer. *Dis Markers* (2017) 2017:7243968. doi: 10.1155/2017/7243968
14. Jiang X, Zhang G, Wu J, Lin S, Liu Y, Ma Y, et al. Long Noncoding RNA Serve as a Potential Predictive Biomarker for Breast Cancer: A Meta-Analysis. *BioMed Res Int* (2020) 2020:9045786. doi: 10.1155/2020/9045786
15. Volovat SR, Volovat C, Hordila I, Hordila DA, Mirestean CC, Miron OT, et al. MiRNA and LncRNA as Potential Biomarkers in Triple-Negative Breast Cancer: A Review. *Front Oncol* (2020) 10:526850. doi: 10.3389/fonc. 2020.526850
16. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: A Powerful Link Between Biological Databases and Microarray Data Analysis. *Bioinformatics* (2005) 21(16):3439–40. doi: 10.1093/bioinformatics/bti525
17. Gu Z, Eils R, Schlesner M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* (2016) 32 (18):2847–9. doi: 10.1093/bioinformatics/btw313
18. Best DJ, Roberts DE. Algorithm AS 89: The Upper Tail Probabilities of Spearman's Rho. *Appl Stat* (1975) 24(3):377. doi: 10.2307/2347111
19. Bolar K. *Stat: Interactive Document for Working With Basic Statistical Analysis. R Package Version 0.1.0* (2019). Available at: https://CRAN.R- project.org/package=STAT.
20. Therneau T. *A Package for Survival Analysis in R_. R Package Version 3.2-10*
(2021). Available at: https://CRAN.R-project.org/package=survival.
21. Korotkevich G, Sukhov V, Sergushichev A. Fast Gene Set Enrichment Analysis. *bioRxiv* (2019) 060012. doi: 10.1101/060012
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
23. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov Jill P, Tamayo P.The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst*
(2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
24. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, et al. Prognostic and Predictive Value of Tumor-Infiltrating Lymphocytes in a Phase III Randomized Adjuvant Breast Cancer Trial in Node-Positive Breast Cancer Comparing the Addition of Docetaxel to Doxorubicin With Doxorubicin-Based Chemotherapy: BIG 02-98. *J Clin Oncol* (2013) 31(7):860–7. doi: 10.1200/JCO.2011.41.0902
25. Ma B, Liao T, Wen D, Dong C, Zhou L, Yang S, et al. Long Intergenic non- Coding RNA 271 is Predictive of a Poorer Prognosis of Papillary Thyroid Cancer. *Sci Rep* (2016) 6:36973. doi: 10.1038/srep36973
26. Buishand FO, Liu-Chittenden Y, Fan Y, Tirosh A, Gara SK, Patel D, et al. Adrenocortical Tumors Have a Distinct, Long, non-Coding RNA Expression Profile and LINC00271 is Downregulated in Malignancy. *Surgery* (2020) 167 (1):224–32. doi: 10.1016/j.surg.2019.04.067
27. Fu X, Zhang L, Dan L, Wang K, Xu Y. LncRNA EWSAT1 Promotes Ovarian Cancer Progression Through Targeting miR-330-5p Expression. *Am J Transl Res* (2017) 9(9):4094–103.
28. Zhou Q, Xie Y, Wang L, Xu T, Gao Y. LncRNA EWSAT1 Upregulates CPEB4 Via miR-330-5p to Promote Cervical Cancer Developm*Mol Cell Biochem* (2020) 471(1-2):177–88. doi: 10.1007/s11010-020- 03778-8
29. Liu J, Huang S, Liao X, Chen Z, Li L, Yu L, et al. LncRNA EWSAT1 Promotes Colorectal Cancer Progression Through Sponging miR-326 to

Modulate FBXL20 Expression. *Onco Targets Ther* (2021) 14:367–78. doi: 10.2147/ OTT.S272895
30. Xu D, Dong P, Xiong Y, Chen R, Konno Y, Ihira K, et al. PD-L1 Is a Tumor Suppressor in Aggressive Endometrial Cancer Cells and Its Expression Is Regulated by miR-216a and lncRNA Meg3. *Front Cell Dev Biol* (2020) 8:598205. doi: 10.3389/fcell.2020.598205
31. Ghafouri-Fard S, Taheri M. Maternally Expressed Gene 3 (MEG3): A Tumor Suppressor Long non Coding RNA. *BioMed Pharmacother* (2019) 118:109129. doi: 10.1016/j.biopha.2019.10912
32. Al-Rugeebah A, Alanazi M, Parine NR. MEG3: An Oncogenic Long Non-Coding RNA in Different Cancers. *Pathol Oncol Res* (2019) 25(3):859–74. doi: 10.1007/s12253-019-00614-3
33. Zhang LH, Li LH, Zhang PF, Cai YF, Hua D. LINC00957 Acted as Prognostic Marker Was Associated With Fluorouracil Resistance in Human Colorectal Cancer. *Front Oncol* (2019) 9:776. doi: 10.3389/fonc.2019.00776
34. Zhang M, Liu Y, Kong D. Identifying Biomolecules and Constructing a Prognostic Risk Prediction Model for Recurrence in Osteosarcoma. *J Bone Oncol* (2020) 26:100331. doi: 10.1016/j.jbo.2020.100331
35. Lv Y, Wei W, Huang Z, Chen Z, Fang Y, Pan L, et al. Long Non-Coding RNA Expression Profile can Predict Early Recurrence in Hepatocellular Carcinoma After Curative Resection. *Hepatol Res* (2018) 48(13):1140–8. doi: 10.1111/hepr.13220
36. Zhang YP, Cheng YB, Li S, Zhao N, Zhu ZH. An Epithelial-Mesenchymal Transition-Related Long non-Coding RNA Signature to Predict Overall Survival and Immune Microenvironment in Kidney Renal Clear Cell Carcinoma. *Bioengineered* (2021) 12(1):555–64. doi: 10.1080/21655979.2021.1880718
37. Li H, Wang M, Zhou H, Lu S, Zhang B. Long Noncoding RNA EBLN3P Promotes the Progression of Liver Cancer Via Alteration of microRNA-144-3p/DOCK4 Signal. *Cancer Manag Res* (2020) 12:9339–49. doi: 10.2147/cmar.s261976
38. Dai S, Li N, Zhou M, Yuan Y, Yue D, Li T, et al. LncRNA EBLN3P Promotes the Progression of Osteosarcoma Through Modifying the miR-224-5p/Rab10 Signaling Axis. *Sci Rep* (2021) 11(1):1992. doi: 10.1038/s41598-021-81641-6
39. Zarzynska JM. Two Faces of TGF-beta1 in Breast Cancer. *Mediators Inflammation* (2014) 2014:141747. doi: 10.1155/2014/141747
40. De Angelis C, Fu X, Cataldo ML, Nardone A, Pereira R, Veeraraghavan J, et al. Activation of the IFN Signaling Pathway is Associated With Resistance to CDK4/6 Inhibitors and Immune Checkpoint Activation in ER-Positive Breast Cancer. *Clin Cancer Res* (2021). doi: 10.1158/1078-0432.CCR-19-4191
41. Ju Q, Zhao YJ, Ma S, Li XM, Zhang H, Zhang SQ, et al. Genome-Wide Analysis of Prognostic-Related lncRNAs, miRNAs and mRNAs Forming a Competing Endogenous RNA Network in Lung Squamous Cell Carcinoma. *J Cancer Res Clin Oncol* (2020) 146(7):1711–23. doi: 10.1007/s00432-020-03224-8
42. Guo X, Zhang M, Li Q, Zhao J, Wang B, Wang J, et al. Evaluation of Genetic Variants in MIR3142HG in Susceptibility to and Prognosis of Glioma. *Am J Clin Oncol* (2020) 43(1):1–8. doi: 10.1097/COC.0000000000000587
43. Hadjicharalambous MR, Roux BT, Feghali-Bostwick CA, Murray LA, Clarke DL, Lindsay MA. Long Non-Coding RNAs Are Central Regulators of the IL-1β- Induced Inflammatory Response in Normal and Idiopathic Pulmonary Lung Fibroblasts. *Front Immunol* (2018) 9:2906. doi: 10.3389/fimmu.2018.02906
44. Ma W, Zhao F, Yu X, Guan S, Suo H, Tao Z, et al. Immune-Related lncRNAs as Predictors of Survival in Breast Cancer: A Prognostic Signature. *J Transl Med* (2020) 18(1):442. doi: 10.1186/s12967-020-02522-6
45. Chen Q, Hu L, Huang D, Chen K, Qiu X, Qiu B. Six-LncRNA Immune Prognostic Signature for Cervical Cancer. *Front Genet* (2020) 11:533628. doi: 10.3389/fgene.2020.533628
46. He Y, Wang X. Identification of Molecular Features Correlating With Tumor Immunity in Gastric Cancer by Multi-Omics Data Analysis. *Ann Transl Med* (2020) 8(17):1050. doi: 10.21037/atm-20-922
47. Baker K, Lachapelle J, Zlobec I, Bismar TA, Terracciano L, Foulkes WD. Prognostic Significance of CD8+ T Lymphocytes in Breast Cancer Depends Upon Both Oestrogen Receptor Status and Histological Grade. *Histopathology* (2011) 58(7):1107–16. doi: 10.1111/j.1365-2559.2011.03846.x
48. Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, et al. CD8 T Cell Score as a Prognostic Biomarker for Triple Negative Breast Cancer. *Int J Mol Sci* (2020) 21(18):6968. doi: 10.3390/ijms21186968
49. Li X, Gruosso T, Zuo D, Omeroglu A, Meterissian S, Guiot MC, et al. Infiltration of CD8+ T Cells Into Tumor Cell Clusters in Triple-Negative Breast Cancer. *Proc Natl Acad Sci USA* (2019) 116(9):3678–87. doi: 10.1073/ pnas.1817652116
50. Heimes AS, Härtner F, Almstedt K, Krajnak S, Lebrecht A, Battista MJ, et al. Prognostic Significance of Interferon-g and Its Signaling Pathway in Early Breast Cancer Depends on the Molecular Subtypes. *Int J Mol Sci* (2020) 21 (19):7178. doi: 10.3390/ijms21197178
51. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N Engl J Med* (2018) 379(22):2108–21. doi: 10.1056/NEJMoa1809615
52. Yang B, Chou J, Tao Y, Wu D, Wu X, Li X, et al. An Assessment of Prognostic Immunity Markers in Breast Cancer. *NPJ Breast Cancer* (2018) 4:35. doi: 10.1038/s41523-018-0088-0
53. Zhang X, Zhang J, Zhao W, Dong X, Xin P, Liu X, et al. Long non-Coding RNA LINC02446 Suppresses the Proliferation and Metastasis of Bladder Cancer Cells by Binding With EIF3G and Regulating the mTOR Signalling Pathway. *Cancer Gene Ther* (2021). doi: 10.1038/s41417-020-00285-2
54. Tong H, Li T, Gao S, Yin H, Cao H, He W. An Epithelial-Mesenchymal Transition-Related Long Noncoding RNA Signature Correlates With the Prognosis and Progression in Patients With Bladder Cancer. *Biosci Rep* (2021) 41(1):BSR20203944. doi: 10.1042/BSR20203944
55. Meng C, Zhou JQ, Liao YS. Autophagy-Related Long Non-Coding RNA Signature for Ovarian Cancer. *J Int Med Res* (2020) 48(11):300060520970761. doi: 10.1177/0300060520970761
56. Chen P, Gao Y, Ouyang S, Wei L, Zhou M, You H, et al. A Prognostic Model Based on Immune-Related Long Non-Coding RNAs for Patients With Cervical Cancer. *Front Pharmacol* (2020) 11:585255. doi: 10.3389/ fphar.2020.585255
57. Gao M, Wang X, Han D, Lu E, Zhang J, Zhang C, et al. A Six-lncRNA Signature for Immunophenotype Prediction of Glioblastoma Multiforme. *Front Genet* (2021) 11:604655. doi: 10.3389/fgene.2020.604655
58. Zhong Z, Hong M, Chen X, Xi Y, Xu Y, Kong D, et al. Transcriptome Analysis Reveals the Link Between lncRNA-mRNA Co-Expression Network and Tumor Immune Microenvironment and Overall Survival in Head and Neck Squamous Cell Carcinoma. *BMC Med Genomics* (2020) 13(1):57. doi: 10.1186/s12920-020-0707-0
59. Chen GR, Sun W, Zheng K, Zhu W. LINC01857 Promotes the Development of Gastric Cancer by Regulating MicroRNA-200b. *Eur Rev Med Pharmacol Sci* (2020) 24(7):3648–56. doi: 10.26355/eurrev_202004_20827
60. Hu G, Liu N, Wang H, Wang Y, Guo Z. LncRNA LINC01857 Promotes Growth, Migration, and Invasion of Glioma by Modulating miR-1281/TRIM65 Axis. *J Cell Physiol* (2019) 234(12):22009–16. doi: 10.1002/ jcp.28763
61. Xiong Y, Gu Y, Wang F, Li L, Zhu M, Wang N, et al. LINC01857 as an Oncogene Regulates CREB1 Activation by Interacting With CREBBP in Breast Cancer. *J Cell Physiol* (2019) 234(8):14031–9. doi: 10.1002/jcp.28090
62. Li Q, Li B, Lu CL, Wang JY, Gao M, Gao W. LncRNA LINC01857 Promotes Cell Growth and Diminishes Apoptosis Via PI3K/mTOR Pathway and EMT Process by Regulating miR-141-3p/MAP4K4 Axis in Diffuse Large B-cell Lymphoma. *Cancer Gene Ther* (2020). doi: 10.1038/ s41417-020-00267-4
63. Wang Q, Yang Y, Fu X, Wang Z, Liu Y, Li M, et al. Long Noncoding RNA Xxylt1-AS2 Regulates Proliferation and Adhesion by Targeting the RNA Binding Protein FUS in HUVEC. *Atherosclerosis* (2020) 298:58–69. doi: 10.1016/j.atherosclerosis.2020.02.018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9. CONSIDERAÇÕES FINAIS

A análise de lncRNAs em diferentes subtipos do câncer de mama apresentou resultados bastante significativos. Como discutido no capítulo I deste trabalho, os lncRNAs foram altamente subtipos específicos, ressaltando a relevância de melhor compreender a biologia destas moléculas para seu uso como marcadores do câncer de mama.

O estudo da biologia de lncRNAs é ainda bastante inicial, e muitas ferramentas bioinformáticas estão sendo desenvolvidas com este propósito. No capítulo II, apresentamos a ferramenta que foi utilizada neste trabalho para predição de redes de co-expressão mediadas por lncRNAs. O uso de redes de co-expressão é bastante informativo, uma vez que se utiliza na sua predição, informações de expressão gênica também de mRNAs. A biologia dos mRNAs é melhor conhecida, e pode fornecer evidências do mecanismo funcional do lncRNA no fenótipo estudado. O método descrito neste capítulo foi capaz de identificar redes de co-expressão diferencialmente ativas/reprimidas dentro dos subtipos moleculares do câncer de mama. Entre estas redes, identificamos o LINC00504, que apresentou a rede de co-expressão com maior diferença de atividade entre os subtipos mais contrastantes do câncer de mama; além disso, apresentou em sua rede genes já descritos como relevantes no processo de carcinogênese mamária, como *ESR1*, *GATA3* e *FOXA1*.

A fim de buscar por mecanismos funcionais em que o LINC00504 está envolvido no câncer de mama, realizou-se o silenciamento deste lncRNA em linhagens celulares do subtipo luminal A. Este subtipo apresentou maior expressão do LINC00504 e por isso foi escolhido para os ensaios funcionais. Observou-se que o LINC00504 apresenta papel oncogênico no subtipo luminal A, uma vez que o seu silenciamento resultou na diminuição da viabilidade celular e da capacidade clonogênica. Desta forma, pode-se dizer que o método de busca por redes de co-expressão mediada por lncRNAs nos diferentes subtipos do câncer de mama, apresentou resultados bastante promissores também como ferramenta para identificar lncRNAs com relevância funcional no desenvolvimento da doença. Nesta busca, gerou-se uma grande quantidade de dados, disponíveis no material suplementar do artigo, que pode servir como ponto de partida para diversos estudos com este mesmo objetivo. A metodologia bioinformática empregada também foi disponibilizada no material suplementar, e pode ser reproduzida no fenótipo de escolha do pesquisador.

Buscamos adicionalmente por lncRNAs envolvidos em um dos "*Hallmarks*" do câncer que vem ganhando bastante destaque nos últimos anos: a evasão do sistema imune. Os lncRNAs já foram descritos como importantes reguladores da resposta imune tumoral, entretanto, não se conhecia, em detalhes, lncRNAs que estivessem relacionados com a resposta imune nos diferentes subtipos do câncer de mama. Para isto, conforme descrito no capítulo III, utilizou-se dados previamente publicados por Thorsson e colaboradores (2018) em que foi realizada uma

caracterização de subtipos imunes de diversos tipos de câncer, incluindo o de mama. Neste trabalho, entretanto, os autores não avaliaram lncRNAs envolvidos com estes subtipos, e devido à relevância destas moléculas como reguladoras da resposta imune e da alta especificidade dos lncRNAs, foi proposta uma metodologia bioinformática para identificar estes RNAs nos subtipos imunes específicos de cada subtipo molecular.

Ao utilizarmos a metodologia proposta, foi possível identificar lncRNAs que apresentaram relevância na resposta imune nos diferentes subtipos moleculares do câncer de mama. Foram identificados lncRNAs com impacto no prognóstico das pacientes, além disso, foi proposto a caracterização funcional de um lncRNA específico de cada subtipo. Neste trabalho, foi gerada uma grande quantidade de dados, todos disponíveis no material suplementar, que podem ser utilizados como base para pesquisas futuras. Além disso, também está disponível para reprodução a metodologia empregada.

Por fim, considerando o que foi apresentado e discutido nos capítulos I, II e III deste trabalho, pode-se concluir sobre a relevância da melhor compreensão de lncRNAs envolvidos nos subtipos moleculares do câncer de mama. Os lncRNAs apresentaram alta especificidade, o que destaca seu uso como marcadores de prognóstico, e além disso, como elementos chave na regulação dos processos de carcinogênese. O desenvolvimento de ferramentas de bioinformática é de grande importância para estudo e melhor entendimento dos lncRNAs em condições fisiológicas e patológicas. As metodologias propostas neste trabalho, foram bastante eficientes em identificar lncRNAs com relevância funcional no câncer de mama, e poderão ser empregadas no futuro por pesquisadores de diversas áreas.

10. CONCLUSÃO

- Os lncRNAs apresentam perfil de expressão diferencial entre os subtipos do câncer de mama;

- As redes de co-expressão mediada por lncRNAs apresentaram perfil de ativação diferencial nos subtipos moleculares do câncer de mama;

- A partir de dados de análise de dados do TCGA observou-se que LINC00504 apresenta expressão aumentada em pacientes do subtipo luminal A. Este mesmo perfil foi confirmado em uma amostra de pacientes brasileiras diagnosticadas com este subtipo;

- A rede de co-expressão baseada no LINC00504 envolve genes já sabidamente relevantes no desenvolvimento do câncer de mama, como *ESR1*, *GATA3* e *FOXA1*;

- O LINC00504 apresenta papel oncogênico em linhagens celulares luminal A, uma vez que seu silenciamento resultou em diminuição da viabilidade celular e da capacidade clonogênica;

- Os lncRNAs apresentam perfil de expressão diferencial entre diferentes subtipos imunes relacionados ao câncer de mama;

- A utilização da metodologia proposta na busca por lncRNAs relacionados com os diferentes subtipos imunes foi bem-sucedida e pode ser empregada em futuros estudos.

REFERÊNCIAS

ALCAZAR, Carlos R. Gil del *et al*. Immune Escape in Breast Cancer During In Situ to Invasive Carcinoma Transition. **Cancer Discovery**,v. 7, n. 10, p. 1098-1115, 26 jun. 2017. American Association for Cancer Research (AACR). http://dx.doi.org/10.1158/2159-8290.cd-17-0222

AZAM, Sikandar *et al*. Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. **Rna Biology**, v. 16, n. 8, p. 1001-1009, 4 jun. 2019. Informa UK Limited. http://dx.doi.org/10.1080/15476286.2019.1620061.

BÄR, Christian; CHATTERJEE, Shambhabi; THUM, Thomas. Long Noncoding RNAs in Cardiovascular Pathology, Diagnosis, and Therapy. **Circulation**, v. 134, n. 19, p. 1484-1499, 8 nov. 2016. Ovid Technologies (Wolters Kluwer Health). http://dx.doi.org/10.1161/circulationaha.116.023686

BALAS, Maggie M.; JOHNSON, Aaron M.. Exploring the mechanisms behind long noncoding RNAs and cancer. **Non-Coding Rna Research**, [S.L.], v. 3, n. 3, p. 108-117, set. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.ncrna.2018.03.001

BATEMAN, Adrian C.; SHAW, Emily C.. Breast pathology. **Surgery (Oxford)**, v. 34, n. 1, p. 1-7, jan. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.mpsur.2015.10.002.

BRITT, Kara L.; CUZICK, Jack; PHILLIPS, Kelly-Anne. Key steps for effective breast cancer prevention. **Nature Reviews Cancer**, [S.L.], v. 20, n. 8, p. 417-436, 11 jun. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/s41568-020-0266-x.

BUNCH, Heeyoun. Gene regulation of mammalian long non-coding RNA. **Molecular Genetics And Genomics**, v. 293, n. 1, p. 1-15, 11 set. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00438-017-1370-9.

CASTRO, Mauro A *et al*. Regulators of genetic risk of breast cancer identified by integrative network analysis. **Nature Genetics**, v. 48, n. 1, p. 12-21, 30 nov. 2015. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/ng.3458.

CIPOLLA, Gabriel *et al*. Long Non-Coding RNAs in Multifactorial Diseases: another layer of complexity. **Non-Coding Rna**, v. 4, n. 2, p. 13, 11 maio 2018. MDPI AG. http://dx.doi.org/10.3390/ncrna4020013.

CLARK, Michael B.; MATTICK, John S.. Long noncoding RNAs in cell biology. **Seminars In Cell & Developmental Biology**, v. 22, n. 4, p. 366-376, jun. 2011. Elsevier BV. http://dx.doi.org/10.1016/j.semcdb.2011.01.001

DAI, X.; LI, T.; BAI, Z.; et al. Breast cancer intrinsic subtype classification, clinical use and future trends. **Am J Cancer Res**, v. 5, n. 10, p. 2929–2943, 2015. Disponível em: <www.ajcr.us>.

DENOIX PF. Tumor, Node and Metastasis (TNM). **Bull Inst. Nat Hyg** (Paris) 7:743, 1952

DERRIEN, T *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. **Genome Res** 2012; 22:1775- 89; PMID:22955988; http://dx.doi.org/10.1101/ gr.132159.111.

DIAMANTOPOULOS, Marios A.; TSIAKANIKAS, Panagiotis; SCORILAS, Andreas. Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer. Annals Of Translational Medicine, v. 6, n. 12, p. 241-241, jun. 2018. AME Publishing Company. http://dx.doi.org/10.21037/atm.2018.06.10.

DURET, L. *et al*. The Xist RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. **Science**, v. 312, n. 5780, p. 1653-1655, 16 jun. 2006. American Association for the Advancement of Science (AAAS). http://dx.doi.org/10.1126/science.1126316

DURINCK, S. *et al*. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. **Bioinformatics**, v. 21, n. 16, p. 3439-3440, 4 ago. 2005. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/bti525

ELISAPHENKO, Eugeny A. *et al*. A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. **Plos One**, v. 3, n. 6, p. 2521-2531, 25 jun. 2008. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pone.0002521

EROLES, Pilar *et al*. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. **Cancer Treatment Reviews**, v. 38, n. 6, p. 698-707, out. 2012. Elsevier BV. http://dx.doi.org/10.1016/j.ctrv.2011.11.005.

FAGHIHI, Mohammad. *et al*. Evidence for natural antisense transcript-mediated inhibition of microRNA function. **Genome Biology**, v. 11, n. 5, p. 1-10, May 2010. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/gb-2010-11-5-r56.

GARCIA,L *et al.* lncRNAs in Hallmarks of Cancer and Clinical Applications. **IntechOpen**. DOI: 10.5772/intechopen.88903. 2019

GEISLER, Sarah; COLLER, Jeff. RNA in unexpected places: long non-coding rna functions in diverse cellular contexts. **Nature Reviews Molecular Cell Biology**, v. 14, n. 11, p. 699-712, 9 out. 2013. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/nrm3679.

GIL, Noa; ULITSKY, Igor. Regulation of gene expression by cis-acting long non-coding RNAs. **Nature Reviews Genetics**, v. 21, n. 2, p. 102-117, 15 nov. 2019. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/s41576-019-0184-5.

GOLDHIRSCH, A. *et al*. Personalizing the treatment of women with early breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2013. **Annals Of Oncology**, v. 24, n. 9, p. 2206-2223, set. 2013. Elsevier BV. http://dx.doi.org/10.1093/annonc/mdt303

GOOD, M. C.; ZALATAN, J. G.; LIM, W. A.. Scaffold Proteins: hubs for controlling the flow of cellular information. **Science**, v. 332, n. 6030, p. 680-686, 5 maio 2011. American Association for the Advancement of Science (AAAS). http://dx.doi.org/10.1126/science.1198701

GU, Zuguang; EILS, Roland; SCHLESNER, Matthias. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847-2849, 20 maio 2016. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/btw313

GU, Zuguang; EILS, Roland; SCHLESNER, Matthias. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847-2849, 20 maio 2016. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/btw313.

GUO, Chun-Jie *et al*. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. **Cell**, v. 181, n. 3, p. 621-636, abr. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2020.03.006.

HANAHAN, Douglas; A WEINBERG, Robert. The Hallmarks of Cancer. **Cell**, v. 100, n. 1, p. 57-70, jan. 2000. Elsevier BV. http://dx.doi.org/10.1016/s0092-8674(00)81683-9

HANAHAN, Douglas; WEINBERG, Robert A.. Hallmarks of Cancer: the next generation. **Cell**, v. 144, n. 5, p. 646-674, mar. 2011. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2011.02.013

HENG, Yujing J *et al*. The molecular basis of breast cancer pathological phenotypes. **The Journal Of Pathology**, v. 241, n. 3, p. 375-391, 29 dez. 2016. Wiley. http://dx.doi.org/10.1002/path.4847

HUARTE, Maite *et al*. A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. **Cell**, v. 142, n. 3, p. 409-419, ago. 2010. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2010.06.040

HUNG, Tiffany *et al*. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. **Nature Genetics**, v. 43, n. 7, p. 621-629, 5 jun. 2011. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/ng.848

INSTITUTO NACIONAL DE CANCER JOSÉ ALENCAR GOMES DA SILVA. **INCA - Instituto Nacional de Câncer - Estimativa 2020**. 2020.

KANG, Mk *et al*. Epigenetic gene regulation by histone demethylases: emerging role in oncogenesis and inflammation. **Oral Diseases**, v. 23, n. 6, p. 709-720, 15 set. 2016. Wiley. http://dx.doi.org/10.1111/odi.12569.

KARAKAS, Didem; OZPOLAT, Bulent. The Role of LncRNAs in Translation. **Non-Coding Rna**, v. 7, n. 1, p. 16, 20 fev. 2021. MDPI AG. http://dx.doi.org/10.3390/ncrna7010016.

KIM, Tae-Kyung *et al*. Enhancer RNAs: a class of long noncoding rnas synthesized at enhancers. **Cold Spring Harbor Perspectives In Biology**, v. 7, n. 1, p. 018622-018644, jan. 2015. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/cshperspect.a018622

KOROTKEVICH, Gennady; SUKHOV, Vladimir; BUDIN, Nikolay; SHPAK, Boris; ARTYOMOV, Maxim N.; SERGUSHICHEV, Alexey. Fast gene set enrichment analysis. **Biorxiv**, p. 1-10, 20 jun. 2016. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/060012

KUNDAJE, A. *et al*. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. **Genome Research**, v. 22, n. 9, p. 1735-1747, 1 set. 2012. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/gr.136366.111

KUO, Chao-Chung *et al*. Detection of RNA–DNA binding sites in long noncoding RNAs. **Nucleic Acids Research**, v. 47, n. 6, p. 32-52, 30 jan. 2019. Oxford University Press (OUP). http://dx.doi.org/10.1093/nar/gkz037

LANDER, E s *et al*. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, 15 fev. 2001. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/35057062

LIBERZON, Arthur *et al*. The Molecular Signatures Database Hallmark Gene Set Collection. **Cell Systems**, v. 1, n. 6, p. 417-425, dez. 2015. Elsevier BV. http://dx.doi.org/10.1016/j.cels.2015.12.004

LIU, Minghui *et al*. Long noncoding RNAs in renal diseases. **Exrna**, v. 1, n. 1, p. 27-50, dez. 2019. AME Publishing Company. http://dx.doi.org/10.1186/s41544-019-0033-x

LONG, Yicheng; WANG, Xueyin; YOUMANS, Daniel T.; CECH, Thomas R.. How do lncRNAs regulate transcription? **Science Advances**, v. 3, n. 9, p. 2110-2120, set. 2017. American Association for the Advancement of Science (AAAS). http://dx.doi.org/10.1126/sciadv.aao2110.

LYNN, Rachel Litman *et al*. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. **Nature**, v. 471, n. 7339, p. 532-536, mar. 2011. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/nature09772

MA, Lina *et al*. On the classification of long non-coding RNAs. **Rna Biology**, v. 10, n. 6, p. 924-933, jun. 2013. Informa UK Limited. http://dx.doi.org/10.4161/rna.24604.

MAKKI, Jaafar. Diversity of Breast Carcinoma: histological subtypes and clinical relevance. **Clinical Medicine Insights**: Pathology, v. 8, p. 31563-31573, jan. 2015. SAGE Publications. http://dx.doi.org/10.4137/cpath.s31563

MATTICK, John s *et al*. Non-coding RNAs: the architects of eukaryotic complexity. **Embo Reports**, v. 2, n. 11, p. 986-991, nov. 2001. EMBO. http://dx.doi.org/10.1093/embo-reports/kve230.

MATTIOLI, Kaia *et al*. High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. **Genome Research**, v. 29, n. 3, p. 344-355, 25 jan. 2019. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/gr.242222.118

NIELSEN, Torsten O. *et al*. A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast Cancer. **Clinical Cancer Research**, v. 16, n. 21, p. 5222-5232, 13 set. 2010. American Association for Cancer Research (AACR). http://dx.doi.org/10.1158/1078-0432.ccr-10-1282.

OLIVEIRA, Jaqueline Carvalho *et al*. LncRNAs in Cancer: another layer of complexity. **The Journal Of Gene Medicine**, p. 3065-3075, 14 dez. 2018. Wiley. http://dx.doi.org/10.1002/jgm.3065

ØROM, Ulf Andersson *et al*. Long Noncoding RNAs with Enhancer-like Function in Human Cells. **Cell**, v. 143, n. 1, p. 46-58, out. 2010. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2010.09.001

PEROU, Charles M. *et al*. Molecular portraits of human breast tumours. **Nature**, v. 406, n. 6797, p. 747-752, ago. 2000. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/35021093

PONTING, Chris P. *et al*. Evolution and Functions of Long Noncoding RNAs. **Cell**, v. 136, n. 4, p. 629-641, fev. 2009. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2009.02.006.

PRAT, Aleix *et al*. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. **Breast Cancer Research**, v. 12, n. 5, p. 1-10, 2 set. 2010. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/bcr2635

QUINN, Jeffrey J.; CHANG, Howard Y.. Unique features of long non-coding RNA biogenesis and function. **Nature Reviews Genetics**, v. 17, n. 1, p. 47-62, 15 dez. 2015. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/nrg.2015.10

RASHID, Farooq; SHAH, Abdullah; SHAN, Ge. Long Non-coding RNAs in the Cytoplasm. **Genomics, Proteomics & Bioinformatics**, v. 14, n. 2, p. 73-80, abr. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.gpb.2016.03.005

RINN, John L.; CHANG, Howard Y.. Genome Regulation by Long Noncoding RNAs. **Annual Review Of Biochemistry**, v. 81, n. 1, p. 145-166, 7 jul. 2012. Annual Reviews. http://dx.doi.org/10.1146/annurev-biochem-051410-092902

SADEGHALVAD, Mona *et al*. Immune microenvironment in different molecular subtypes of ductal breast carcinoma. **Breast Cancer Research And Treatment**, v. 185, n. 2, p. 261-279, 3 out. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s10549-020-05954-2.

SALDAÑA-MEYER, Ricardo *et al*. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. **Molecular Cell**, v. 76, n. 3, p. 412-422, nov. 2019. Elsevier BV. http://dx.doi.org/10.1016/j.molcel.2019.08.015

SALMENA, Leonardo *et al*. A ceRNA Hypothesis: the rosetta stone of a hidden rna language?. **Cell**, v. 146, n. 3, p. 353-358, ago. 2011. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2011.07.014

SANGA, Sandeep *et al*. Gene expression meta-analysis supports existence of molecular apocrine breast cancer with a role for androgen receptor and implies interactions with ErbB family. **Bmc Medical Genomics**,v. 2, n. 1, p. 1-10, 11 set. 2009. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/1755-8794-2-59

SØRLIE, Therese *et al*. Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings Of The National Academy Of Sciences**, v. 100, n. 14, p. 8418-8423, 26 jun. 2003. Proceedings of the National Academy of Sciences. http://dx.doi.org/10.1073/pnas.0932692100

SUBRAMANIAN, A. *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. **Proceedings Of The National Academy Of Sciences**, v. 102, n. 43, p. 15545-15550, 30 set. 2005. Proceedings of the National Academy of Sciences. http://dx.doi.org/10.1073/pnas.0506580102

SUGITA, Bruna *et al*. Differentially expressed miRNAs in triple negative breast cancer between African-American and non-Hispanic white women. **Oncotarget**, v. 7, n. 48, p. 79274-79291, 2 nov. 2016. Impact Journals, LLC. http://dx.doi.org/10.18632/oncotarget.13024.

SUNG, Hyuna; FERLAY, Jacques; SIEGEL, Rebecca L.; LAVERSANNE, Mathieu; SOERJOMATARAM, Isabelle; JEMAL, Ahmedin; BRAY, Freddie. Global Cancer Statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **Ca**: A Cancer Journal for Clinicians, v. 71, n. 3, p. 209-249, 4 fev. 2021. Wiley. http://dx.doi.org/10.3322/caac.21660.

TAMIMI, Rulla M *et al*. Comparison of molecular phenotypes of ductal carcinoma in situand invasive breast cancer. **Breast Cancer Research**, v. 10, n. 4, p. 1-10, ago. 2008. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/bcr2128

THERNEAU T. A Package for Survival Analysis in R_. R Package Version 3.2-10 (2021). Available at: https://CRAN.R-project.org/package=survival.

THORSSON, Vésteinn *et al*. The Immune Landscape of Cancer. **Immunity**, v. 48, n. 4, p. 812-830, abr. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.immuni.2018.03.023.

VENTER, J. Craig *et al*. The Sequence of the Human Genome. **Science**, v. 291, n. 5507, p. 1304-1351, 16 fev. 2001. American Association for the Advancement of Science (AAAS). http://dx.doi.org/10.1126/science.1058040

WANG, Kevin C. *et al*. Molecular Mechanisms of Long Noncoding RNAs. **Molecular Cell**, v. 43, n. 6, p. 904-914, set. 2011. Elsevier BV. http://dx.doi.org/10.1016/j.molcel.2011.08.018

WU, Tao; DU, Yantao. LncRNAs: from basic research to medical application. **International Journal Of Biological Sciences**, v. 13, n. 3, p. 295-307, 2017. Ivyspring International Publisher. http://dx.doi.org/10.7150/ijbs.16968

XING, Zhen *et al*. LncRNA Directs Cooperative Epigenetic Regulation Downstream of Chemokine Signals. **Cell**, v. 159, n. 5, p. 1110-1125, nov. 2014. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2014.10.013.

YANG, Liuqing *et al*. LncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. **Nature**, v. 500, n. 7464, p. 598-602, 14 ago. 2013. Springer Science and Business Media LLC. http://dx.doi.org/10.1038/nature12451

YOON, Je-Hyun; ABDELMOHSEN, Kotb; GOROSPE, Myriam. Posttranscriptional Gene Regulation by Long Noncoding RNA. **Journal Of Molecular Biology**, v. 425, n. 19, p. 3723-3730, out. 2013. Elsevier BV. http://dx.doi.org/10.1016/j.jmb.2012.11.024

ZHANG, Xiaopei *et al*. Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. **International Journal Of Molecular Sciences**, v. 20, n. 22, p. 5573, 8 nov. 2019. MDPI AG. http://dx.doi.org/10.3390/ijms20225573.

ZHANG, Xueqing; WEISSMAN, Sherman M; NEWBURGER, Peter e. Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. **Rna Biology**,v. 11, n. 6, p. 777-787, 24 abr. 2014. Informa UK Limited. http://dx.doi.org/10.4161/rna.28828

APÊNDICES

PRODUÇÃO CIENTÍFICA ADICIONAL DESENVOLVIDA NO PERÍODO DE DOUTORADO

APÊNDICE 1: Artigo intitulado "*So alike yet so different. Differential expression of the long non-coding RNAs NORAD and HCG11 in breast cancer subtypes*" publicado em 2021 na revista **Genetics And Molecular Biology**.

APÊNDICE 2: Artigo intitulado "*Comprehensive Analysis of ceRNA Networks in HPV16- and HPV18-mediated Cervical Cancers Reveals XIST as a Pivotal Competing Endogenous RNA*" publicado em 2021 na revista **Biochimica Et Biophysica Acta-Molecular Basis Of Disease**.

APÊNDICE 3: Artigo intitulado "*Association between SNP rs527616 in lncRNA AQP4-AS1 and susceptibility to breast cancer in a southern Brazilian population*" publicado em 2021 na revista **Genetics And Molecular Biology**.

APÊNDICE 4: Artigo intitulado "*Polymorphism of lncRNAs in breast cancer: Meta-analysis shows no association with susceptibility*" publicado em 2020 na revista **The Journal of Gene Medicine**.

APÊNDICE 5: Artigo intitulado "*Highlighting transcribed ultraconserved regions in human diseases*" publicado em 2019 na revista **Wiley Interdisciplinary Reviews-RNA**.

APÊNDICE 6: Artigo intitulado "*A genetic variant in microRNA-146a is associated with sporadic breast cancer in a Southern Brazilian Population*" publicado em 2019 na revista **Genetics And Molecular Biology**.

APÊNDICE 7: Artigo intitulado "*Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity*" publicado em 2018 na revista **Non-Coding RNA**.

APÊNDICE 8: Artigo intitulado "*LncRNAs in Cancer: another layer of complexity*" publicado em 2018 na revista **The Journal of Gene Medicine**.

APÊNDICE 9: Artigo intitulado "*Long Non-Coding RNA TUG1 Expression Is Associated with Different Subtypes in Human Breast Cancer*" publicado em 2017 na revista **Non-Coding RNA**.

APÊNDICE 1: https://doi.org/10.1590/1678-4685-GMB-2020-0153

Genetics and Molecular Biology, 44, 1, e20200153 (2021)

DOI: https://doi.org/10.1590/1678-4685-GMB-2020-0153

Research Article
Human and Medical Genetics

So alike yet so different. Differential expression of the long non-coding RNAs NORAD and HCG11 in breast cancer subtypes

Carolina Mathias*[1], Gabrielle Araújo Pedroso*[1], Fernanda Rezende Pabst[1], Rubens Silveira de Lima[2], Flavia Kuroda[2], Iglenir João Cavalli[1], Jaqueline Carvalho de Oliveira[1], Enilze Maria de Souza Fonseca Ribeiro[1] and Daniela Fiori Gradia[1]

[1]*Universidade Federal do Paraná, Departamento de Genética, Programa de Pós Graduação em Genética, Curitiba, PR, Brazil.*
[2]*Hospital Nossa Senhora das Graças, Centro de Doenças da Mama, Curitiba, PR, Brazil.*
* These authors contributed equally to this work.

APÊNDICE 2: https://doi.org/10.1016/j.bbadis.2021.166172

Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease
Volume 1867, Issue 10, 1 October 2021, 166172

Comprehensive analysis of ceRNA networks in HPV16- and HPV18-mediated cervical cancers reveals XIST as a pivotal competing endogenous RNA

Fernanda Costa Brandão Berti [a, b, 1], Carolina Mathias [a, c, 1], Leandro Encarnação Garcia [a, c], Daniela Fiori Gradia [a, c], Patrícia Savio de Araújo-Souza [a, d], Gabriel Adelman Cipolla [a, b], Jaqueline Carvalho de Oliveira [a, c], Danielle Malheiros [a, b]

APÊNDICE 3: https://doi.org/10.1590/1678-4685-GMB-2020-0216

Genetics and Molecular Biology, 44, 1, e20200216 (2021)

DOI: https://doi.org/10.1590/1678-4685-GMB-2020-0216

Research Article
Human and Medical Genetics

Association between SNP rs527616 in lncRNA *AQP4-AS1* and susceptibility to breast cancer in a southern Brazilian population

Rafael D. Marchi[1], Carolina Mathias[1], Gabriel A. K. Reiter[1], Rubens Silveira de Lima[2], Flávia Kuroda[2], Cícero de Andrade Urban[2], Ricardo. L. R. de Souza[1], Daniela F. Gradia[1], Enilze M. S. F. Ribeiro[1], Iglenir J. Cavalli[1] and Jaqueline Carvalho de Oliveira[1]

[1]*Universidade Federal do Paraná, Departamento de Genética, Curitiba, PR, Brazil.*
[2]*Hospital Nossa Senhora das Graças, Centro de Doenças da Mama, Curitiba, PR, Brazil.*

APÊNDICE 4: https://doi.org/10.1002/jgm.3271

The Journal of Gene Medicine

REVIEW ARTICLE

Polymorphism of lncRNAs in breast cancer: Meta-analysis shows no association with susceptibility

Carolina Mathias, Leandro E. Garcia, Mayza Dalcin Teixeira, Ana Flávia Kohler, Rafael D. Marchi, Jéssica F. Barazetti, Daniela Fiori Gradia, Jaqueline Carvalho de Oliveira

First published: 05 September 2020 | https://doi.org/10.1002/jgm.3271 | Citations: 1

Carolina Mathias and Leandro E. Garcia contributed equally to this work.

APÊNDICE 5: https://doi.org/10.1002/wrna.1567

Received: 27 May 2019 | Revised: 2 August 2019 | Accepted: 13 August 2019
DOI: 10.1002/wrna.1567

ADVANCED REVIEW

WIREs RNA WILEY

Highlighting transcribed ultraconserved regions in human diseases

Erika Pereira Zambalde[1] | Carolina Mathias[1] | Ana Carolina Rodrigues[1] | Enilze M. de Souza Fonseca Ribeiro[1] | Daniela Fiori Gradia[1] | George A. Calin[2] | Jaqueline Carvalho de Oliveira[1]

APÊNDICE 6: https://doi.org/10.1590/1678-4685-GMB-2019-0278

Genetics and Molecular Biology, 42, 4, e20190278 (2019)

DOI: http://dx.doi.org/ 10.1590/1678-4685-GMB-2019-0278

Original Article
Human and Medical Genetics

A genetic variant in microRNA-146a is associated with sporadic breast cancer in a Southern Brazilian Population

Heloisa Magagnin Brincas[1], Danillo G. Augusto[1†], Carolina Mathias[1], Iglenir João Cavalli[1], Rubens Silveira de Lima[2], Flávia Kuroda[2], Cícero de Andrade Urban[2], Daniela Fiori Gradia[1], Jaqueline de Oliveira[1], Rodrigo Coutinho de Almeida[1†] and Enilze Maria de Souza Fonseca Ribeiro[1]

[1]*Universidade Federal do Paraná, Departamento de Genética, Curitiba, Paraná, Brazil.*
[2]*Hospital Nossa Senhora das Graças, Centro de Doenças da Mama, Curitiba, Paraná, Brazil.*

APÊNDICE 7: https://doi.org/10.3390/ncrna4020013

non-coding RNA

MDPI

Review

Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity

Gabriel A. Cipolla, Jaqueline C. de Oliveira, Amanda Salviano-Silva, Sara C. Lobo-Alves, Debora S. Lemos, Luana C. Oliveira, Tayana S. Jucoski, Carolina Mathias, Gabrielle A. Pedroso, Erika P. Zambalde and Daniela F. Gradia *

Department of Genetics, Federal University of Parana, Curitiba 81531-980, Brazil; gabriel.cipolla@ufpr.br (G.A.C.); jaqueline.carvalho@ufpr.br (J.C.d.O.); amandasalviano@ufpr.br (A.S.-S.); sara.alves@ufpr.br (S.C.L.-A.); debora.lemos@ufpr.br (D.S.L.); luana.oliveira@ufpr.br (L.C.O.); tayanajucoski@ufpr.br (T.S.J.); carolina.mathias@ufpr.br (C.M.); gabriellearaujop@ufpr.br (G.A.P.); erikazambalde@ufpr.br (E.P.Z.)
* Correspondence: danielagradia@ufpr.br; Tel.: +55-41-3361-1725

Received: 26 February 2018; Accepted: 4 May 2018; Published: 11 May 2018

check for updates

APÊNDICE 8: https://doi.org/10.1002/jgm.3065

The Journal of Gene Medicine

REVIEW ARTICLE

Long non-coding RNAs in cancer: Another layer of complexity

Jaqueline Carvalho de Oliveira, Luana Caroline Oliveira, Carolina Mathias, Gabrielle Araújo Pedroso, Debora Souza Lemos, Amanda Salviano-Silva, Tayana Schultz Jucoski ... See all authors

First published: 14 December 2018 | https://doi.org/10.1002/jgm.3065 | Citations: 21

APÊNDICE 9: https://doi.org/10.3390/ncrna3040026

non-coding RNA

MDPI

Communication

Long Non-Coding RNA *TUG1* Expression Is Associated with Different Subtypes in Human Breast Cancer

Daniela F. Gradia *,† , Carolina Mathias † , Rodrigo Coutinho, Iglenir J. Cavalli, Enilze M. S. F. Ribeiro and Jaqueline C. de Oliveira

Department of Genetics, Federal University of Parana, Curitiba 81531980, Brazil; danielagradia@gmail.com (D.F.G.); carol.mathias1@hotmail.com (C.M.); rodrigocout@gmail.com (R.C.); cavalli@ufpr.br (I.J.C.); enilzeribeiro@gmail.com (E.M.S.F.R.); jaqueline.genetica@gmail.com (J.C.O.)

* Correspondence: danielagradia@ufpr.br; Tel.: +55-41-3361-1725

† These authors contributed equally to this work.

Received: 24 October 2017; Accepted: 19 December 2017; Published: 20 December 2017